



**QUEEN'S
UNIVERSITY
BELFAST**

DOCTOR OF PHILOSOPHY

Measuring critical thinking skills and dispositions in undergraduate students

O'Hare, Liam

Award date:
2004

Awarding institution:
Queen's University Belfast

[Link to publication](#)

Terms of use

All those accessing thesis content in Queen's University Belfast Research Portal are subject to the following terms and conditions of use

- Copyright is subject to the Copyright, Designs and Patent Act 1988, or as modified by any successor legislation
- Copyright and moral rights for thesis content are retained by the author and/or other copyright owners
- A copy of a thesis may be downloaded for personal non-commercial research/study without the need for permission or charge
- Distribution or reproduction of thesis content in any format is not permitted without the permission of the copyright holder
- When citing this work, full bibliographic details should be supplied, including the author, title, awarding institution and date of thesis

Take down policy

A thesis can be removed from the Research Portal if there has been a breach of copyright, or a similarly robust reason. If you believe this document breaches copyright, or there is sufficient cause to take down, please contact us, citing details. Email: openaccess@qub.ac.uk

Supplementary materials

Where possible, we endeavour to provide supplementary materials to theses. This may include video, audio and other types of files. We endeavour to capture all content and upload as part of the Pure record for each thesis.

Note, it may not be possible in all instances to convert analogue formats to usable digital formats for some supplementary materials. We exercise best efforts on our behalf and, in such instances, encourage the individual to consult the physical thesis for further information.



**QUEEN'S
UNIVERSITY
BELFAST**

Measuring critical thinking skills and dispositions in undergraduate students

O'Hare, L. (2004). Measuring critical thinking skills and dispositions in undergraduate students. Queens University Belfast.

Queen's University Belfast - Research Portal:

[Link to publication record in Queen's University Belfast Research Portal](#)

Publisher rights

Copyright the author 2004.

General rights

Copyright for the publications made accessible via the Queen's University Belfast Research Portal is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

The Research Portal is Queen's institutional repository that provides access to Queen's research output. Every effort has been made to ensure that content in the Research Portal does not infringe any person's rights, or applicable UK laws. If you discover content in the Research Portal that you believe breaches copyright or violates any law, please contact openaccess@qub.ac.uk.

Measuring Critical Thinking Skills and Dispositions in Undergraduate Students

Liam O'Hare BSc.

A thesis submitted to the Faculty of Science and Agriculture
of the Queen's University of Belfast
in fulfilment of the requirement for the degree
of Doctor of Philosophy

School of Psychology

October 2004

Abstract

This thesis investigates critical thinking with a particular focus on measurement in undergraduate students. A higher education context was chosen because many regard critical thinking development as a primary goal for third level education. Nine studies, both cross-sectional and longitudinal in design, were conducted with undergraduate psychology students (N=387), using the California Critical Thinking Disposition Inventory (CCTDI) and the California Critical Thinking Skills Test (CCTST). Studies 1-3 revealed psychometric weaknesses in the CCTDI and revised the scale with factor analysis and reliability analysis to form the CCTDI United Kingdom revision (CCTDI-UK). Study 4 investigated convergent validity and showed significant inter-correlation between the sub-scales of the CCTDI-UK, and significant correlations with the Openness scale of the NEO Personality Inventory (NEO PI-R). The study also provided evidence for improvement in scores on three of the six sub-scales in the CCTDI-UK (Truth-Seeking, Inquisitiveness, Open-Mindedness) during the course of an undergraduate degree. Study 5 explored a two factor structure for critical thinking dispositions. Study 6 used reliability analysis to revise the CCTST to produce the CCTST-UK. Study 7 showed that the CCTST-UK had a moderate correlation with degree attainment and a slightly higher correlation with a test of non-verbal intelligence (Raven's Advanced Progressive Matrices short form); in addition, the study showed that scores on the CCTST-UK improved during the course of the degree. Studies 8 and 9 investigated the potential of critical thinking for predicting degree attainment. A-levels predicted approximately 10% of the variance of degree attainment while entry level scores on the CCTST-UK predicted an additional 5%. Exit level scores on the CCTST-UK and the Inquisitive sub-scale of the CCTDI-UK were found to be predictors of degree attainment. The main conclusions of the thesis were that these tests had significant potential for predicting degree attainment and that they measured a substantial proportion of the theoretical constructs identified by the major authors in critical thinking.

Table of Contents

Acknowledgements.....	i
Abstract.....	iii
List of Tables.....	iv
List of Figures.....	vii

Chapter 1 Critical Thinking: History and Perspectives

1.1 Introduction.....	1
1.2 Critical Thinking: What is it? Why is it important?.....	2
1.3 A History of ‘Critical Thinking’.....	5
1.4 Critical Thinking Theory: Born in Philosophy - Adopted by Psychology.....	11
1.5 Critical Thinking Applications: Education and Beyond.....	15
1.6 Conclusion.....	19

Chapter 2 Critical Thinking: Definitions, Concepts Frameworks and Models

2.1 Introduction.....	21
2.2 Critical Thinking: A Statement of Expert Consensus for Purposes of Educational Assessment and Instruction. ‘The Delphi Report’	23
2.3 Robert Ennis	
2.3.1 Ennis’s Definitions of critical thinking.....	31
2.3.2 Ennis’ Theories on Critical Thinking.....	31
2.3.3 Evaluating Ennis.....	35
2.4 Richard Paul	
2.4.1 Paul’s Definition of Critical Thinking.....	36
2.4.2 Paul’s Models of Critical Thinking.....	38
2.4.3 Evaluating Paul.....	40
2.5 Alec Fisher	
2.5.1 Fisher’s Definition of Critical Thinking.....	40
2.5.2 Fisher’s Model of Critical Thinking Competencies.....	41
2.5.3 Fisher Summary and Conclusion.....	43
2.6 Diane Halpern	
2.6.1 Halpern’s Definition of Critical Thinking.....	44

2.6.2	Halpern's Framework of Critical Thinking.....	45
2.6.3	Evaluating Halpern	47
2.7	Deanna Kuhn	
2.7.1	Kuhn's Definition of Critical Thinking.....	49
2.7.2	Kuhn's Models of Critical Thinking.....	50
2.7.3	Evaluating Kuhn.....	52
2.8	Merging the Models.....	53
2.9	Comparing the Merged Model with the Delphi Analysis.....	56

Chapter 3 The Measurement of Critical Thinking in Higher Education

3.1	Introduction.....	57
3.2	Critical thinking Assessment Issues.....	57
3.2.1	Introduction.....	57
3.2.2	Why is it Critical Thinking Assessment Important in Higher Education?.....	58
3.2.3	Assessing Critical Thinking in Knowledge Domains vs. General Skills.....	61
3.3	Critical Thinking Measurement Tools	
3.3.1	Introduction.....	63
3.3.2	Watson Glaser Critical Thinking Appraisal (WGCTA)...	64
3.3.3	Cornell Critical Thinking Test Level Z (CCTT) (Ennis & Millman 1985).....	67
3.3.4	Ennis Weir Critical Thinking Essay Test; <i>EWCTET</i> (Ennis & Weir 1985).....	69
3.3.5	California Critical Thinking Skills Test (CCTST) Form A.....	71
3.3.6	California Critical Thinking Disposition Inventory (CCTDI).....	74
3.4	Conclusion.....	77

Chapter 4 Rationale and General Design

4.1	Introduction.....	79
4.2	Review of Introductory Chapters.....	80
4.3	Research Questions and Structure of the Empirical Studies.....	81
4.4	General Design.....	85
4.4.1	General Participant Description.....	85
4.4.2	Cohort Descriptions.....	87

Chapter 5 A Psychometric Analysis of Critical Thinking Dispositions: The California Critical Thinking Dispositions Inventory (CCTDI)

5.1	Introduction.....	89
5.2	The Validity of Psychological Measurement.....	89
5.2.1	Introduction.....	89
5.2.1.1	Content Validity.....	90
5.2.1.2	Criterion Validity.....	91
5.2.1.3	Construct Validity.....	93
5.3	Previous Psychometric Examination of the CCTDI.....	95
5.4	Study 1: Internal Consistency and Confirmatory Factor Analysis of the CCTDI.....	97
5.4.1	Introduction.....	97
5.4.2	Method.....	99
5.4.3	Results.....	100
5.4.4	Conclusions.....	101
5.5	Study 2: Exploratory Principle Component and Reliability Analysis of the CCTDI.....	103
5.5.1	Introduction.....	103
5.5.2	Method.....	103
5.5.3	Results - Study 2(a): Exploratory Factor Analysis of the CCTDI.....	104
5.5.4	Results - Study 2(b): Reliability Analysis of the CCTDI-UK.....	120
5.5.5	Conclusions.....	121

5.6	Study 3: Confirmatory Factor Analysis of the CCTDI-UK and Reliability construct validity comparison with the CCTDI.....	122
5.6.1	Introduction.....	122
5.6.2	Method.....	122
5.6.3	Results Study 3(a) - Confirmatory Factor Analysis of the CCTDI-UK and Comparison with the CCTDI.....	123
5.6.4	Study 3(b) - Reliability Comparison between the CCTDI-UK and the CCTDI.....	124
5.6.5	Conclusions.....	124
5.7	Study 4: External Criterion Validity (Convergent and Group Contrast Validity) of the CCTDI-UK.....	125
5.7.1	Study 4(a): Convergent Validity of CCTDI-UK.....	125
5.7.1.1	Introduction.....	125
5.7.1.2	Method.....	126
5.7.1.3	Results.....	127
5.7.1.4	Conclusions.....	128
5.7.2	Study 4(b): Group Contrast Validity of the CCTDI-UK.....	129
5.7.2.1	Introduction.....	129
5.7.2.2	Method.....	129
5.7.2.3	Results.....	131
5.7.2.4	Conclusions.....	134
5.8	Study 5: Investigation of Higher Order Factors of Critical Thinking Dispositions.....	136
5.8.1	Introduction.....	136
5.8.2	Method.....	136
5.8.3	Results	
5.8.4	Conclusions.....	138
5.9	Discussion.....	139

Chapter 6 A Psychometric Analysis of Critical Thinking Skills: The California Critical Thinking Skills Test (CCTST)

6.1	Introduction.....	143
6.2	Study 6: The internal consistency of the CCTST.....	149
6.2.1	Introduction.....	149
6.2.2	Method.....	149
6.2.3	Results.....	150
6.2.4	Conclusions.....	152
6.3	Study 7: Criterion validity of the California Critical Thinking Skills Test-UK.....	153
6.3.1	Introduction.....	153
6.3.2	Method.....	153
6.3.3	Results.....	154
6.4	Discussion.....	158

Chapter 7 Using Scores on the CCTST-UK and CCTDI-UK to Predict Degree Attainment

7.1	Introduction.....	159
7.2	Higher Education Admissions: A Role for Critical Thinking Measurement?.....	160
7.3	Study 8 Using 1 st year Levels of Critical Thinking to Predict Degree Attainment.....	163
7.3.1	Introduction.....	163
7.3.2	Method.....	163
7.3.3	Results.....	164
7.4	Study 9 Using 3 rd Year Levels of Critical Thinking to Predict Degree Attainment.....	170
7.4.1	Introduction.....	170
7.4.2	Method.....	170
7.4.3	Results.....	172
7.5	Discussion.....	176

Chapter 8 General Discussion

8.1	Introduction.....	180
8.2	Synthesis of Experimental Findings.....	181
8.3	Theoretical and Practical Implications.....	189
8.4	Future Directions.....	193
8.5	Conclusion.....	197

References.....	198
------------------------	------------

Appendices for Chapter 2...

Appendix 2.1: Matrix of Major Authors Critical Thinking Models & Frameworks.....	225
---	-----

Appendices for Chapter 3

Appendix 3.1: Matrix of Critical Thinking Test Validity.....	227
Appendix 3.2: Matrix of Critical Thinking Test Reliability.....	229

Appendices for Chapter 5

Appendix 5.1: The California Critical Thinking Disposition Inventory.....	231
Appendix 5.2: Factor Analysis Output (Study 2).....	237

Appendices for Chapter 6

Appendix 5.2: The California Critical Thinking Skills Test.....	239
---	-----

‘Believed, but the seed of doubt was there, and it stayed, and every now and then sent out a little root. It changed everything, to have that seed growing. It made Ender listen more carefully to what people meant, instead of what they said. It made him wise.’

Ender’s Game - Orson Scott Card (1985)

Chapter 1

Critical Thinking: History and Perspectives

1.1 Introduction

The first chapter provides a broad overview of the concept of critical thinking by outlining the traditions of different scholars and writers who claim critical thinking as their area of expertise. It begins by adopting a layperson's perspective and points out why critical thinking is important for everyday reasoning and problem-solving. Next, the chapter reviews the history and philosophy of critical thinking and explains how ideas from ancient Greek philosophers through to early 20th century thinkers are still relevant to the current dialogue on critical thinking. The section also highlights that, throughout history, there have been frequent analyses of what constitutes good critical thinking and many attempts to develop it as a social norm.

The chapter then moves on to discuss contemporary perspectives in critical thinking research. This section describes how critical thinking research has developed from philosophical roots and is currently researched from a number of psychological perspectives. These perspectives include theoretical views of cognitive, developmental, psychometric and social psychologists. The last section identifies recent developments related to critical thinking and higher education in the UK and points to the growing relevance of forms of critical reasoning and thinking across a wide range of professional, business and civic life.

1.2 Critical Thinking: What is it? Why is it important?

What is critical thinking? Not surprisingly, this question is at the heart of the contemporary literature on critical thinking (Beyer, 1985; Ennis, 1990; Fisher, 2001; Fisher and Scriven, 1997; Moran, 1997; van Gelder, 2001). Attempting to answer this question uncovers a wide range of issues. In its simplest form, critical thinking could be said to be about challenging a claim or an opinion (either one's own or another person's) with the purpose of finding out what to believe or do. Indeed, one of the most prolific contemporary writers on critical thinking, Robert Ennis, defined critical thinking in this way.

“Critical thinking is reasonable and reflective thinking that is focused on deciding what to believe or do.”

Norris and Ennis (1989, p.1)

This definition captures the purpose and intention of critical thinking but it does not say much about critical thinking as a cognitive process. However, the following definition by Diane Halpern is more embedded in cognitive theory.

“Critical thinking is the use of those cognitive skills or strategies that increase the probability of a desirable outcome. It is used to describe thinking that is purposeful, reasoned, and goal directed – the kind of thinking involved in solving problems, formulating inferences, calculating likelihood, and making decisions when the thinker is using skills that are thoughtful and effective for the particular context and type of thinking task.”

Halpern (1996, p.5)

Increasingly, the focus has shifted from the process of critical thinking as a cognitive operation to the characteristics of the critical thinker, to the attitudes, inclinations and dispositions of a person who habitually engages in critical thinking and considers it important as a personal, educational and societal value. These attributes are exemplified in the following definition by Richard Paul.

“As we come to think critically in the strong sense we develop special traits of mind: intellectual, humility, intelligence, courage, intellectual perseverance, intellectual integrity and confidence in reason. A sophistic or weak sense critical thinker develops these traits only in a restricted way, consistent with egocentric and socio-economic commitments.”

Richard Paul (1993, p.33)

When critical thinking is considered from a psychological perspective – as a set of cognitive processes and/or thinking dispositions, then some typical psychological questions emerge. What kind of a construct is it? Is it one single coherent construct or are there several components involved? How do the cognitive processes that underpin critical thinking relate to the more dispositional aspects of thinking? How can we measure these constructs? How can they be developed? Does psychology as a discipline adopt different perspectives on the nature of critical thinking? How pervasive is critical thinking in our society? Do we value it? Does our education system promote it? In particular, is it important for learning and teaching in higher education? This thesis attempts to elucidate on some of these central questions, particularly those around the nature of the cognitive processes of critical thinking –

critical thinking skills – and the dispositional aspects of critical thinking – critical thinking dispositions. The main focus is on their assessment.

Why is critical thinking important? Ennis (1996a) suggests that critical thinking is relevant to every part of our private, working and social lives. If people are listening to a party political broadcast, they need to question the validity of the claims being made, the nature of the evidence, the coherence of the argument and so on. Equally, if people are searching through the internet, they need to adopt a critical perspective on the quality of the information being encountered – to what extent is it consistent, are there different points of view being presented, how do these viewpoints arise? It is vital that everybody can think critically about this abundance of views, especially when there is little quality control, as is the case with the Internet (Munger, Anderson, Benjamin, Brusiel, and Paredes-Holt, 1999; Connor-Greene and Greene, 2002). If a person is offered an attractive job in a new country, he/she needs to predict the consequences of the move and weigh up the pros and cons before making a decision. People are constantly bombarded with advertisements making claims about the benefits of all types of products. People need to be able to analyse the nature of the claims and whether they are being persuaded by rhetoric or logic. Ennis (1996a) and Paul (1993) would even argue that we have a responsibility to be critical thinkers in all our civic decisions, such as voting and helping others in civic matters, as this is crucial in maintaining a stable democratic society. Therefore, they would argue, it is increasingly important in a modern world that students receive explicit teaching in critical thinking, to help them make sense of the increasingly diverse views to which they are exposed, to develop standards of evidence and plausibility in arguments, to argue and defend their own positions, to anticipate consequences and

make good personal and civic decisions. Pressely (1990) extends these claims by suggesting that explicit instruction in critical thinking provides a moral education which allows students to process and appraise the diverse views of modern society.

A concern about critical thinking in society is not new. Critical thinking has a long history and has been debated by great minds for over two thousand years (see Section 1.3). Furthermore, there are currently a multitude of disciplines and scholars, ranging in perspectives from pure science to philosophy, which have a theoretical stake in critical thinking. In addition, the practical outcomes of critical thinking (or lack thereof) affect people who range in status from world leaders to those who live in poverty. Section 1.4 focuses on the explicit dialogue around critical thinking currently seen in education at all levels and why critical thinking research promises so much in this area. Lastly, Section 1.4 discusses the uses of critical thinking in a number of professional domains, which shows the influence of this topic in almost every aspect of human society.

1.3 A History of ‘Critical Thinking’

‘The unexamined life is not worth living.’ Socrates (470-399 BC).

The origins of critical thinking can be traced back to antiquity, with the first recorded figure to discuss a critical form of thinking being the Greek philosopher, Socrates.

It is difficult to disentangle the views of Socrates and Plato, as knowledge of both philosophers’ work comes from a few sources, mainly Plato’s Dialogues (Gotlieb, 1999) and Xenophon’s account of Socrates’ trial and death (Tredennick, 1990).

However, a common distinction between these two historical figures is that Plato carried on the work of the pre-Socratics, e.g., Pythagoras, Heraclitus and Parmenides, who tried to produce new solutions to old problems like the nature of reality and the role of God using deductive reasoning. In contrast, Socrates was more interested in Sophist (i.e. ancient Greek equivalent to lawyers) issues which concerned the nature and construction of strong arguments based on particular cases (Mac Donald Ross, 1993). In essence, Socrates used the method of inductive reasoning.

‘The Socratic Method of Philosophy’ has a number of elements that are still visible in the current theories of critical thinking (see Chapter 2). The initial focus is on the ‘Action’, which is to be carried out using the information gained from a dialectically derived truth. Socrates’ core method for discerning this truth is ‘Questioning’, usually performed in a question and answer session. Quite often these sessions involved confronting and challenging someone’s argument. The purpose of the challenge was to find the truth, which is the ultimate goal. Furthermore, Socrates had a personal goal which was to raise the level of human argument, by questioning beliefs and actions, and by increasing human consciousness and self-reflection. These were the characteristics that distinguished humans from other species.

Socratic seminars are now a popular way of introducing Socratic questioning to groups of both children and adults (Fisher, 1990; Frost, 1942; Lipman, 1980; Polite and Adams, 1996). Their function is described in the following quote by Lambright (1995).

'The purpose of Socratic seminars is to enlarge understanding of ideas, issues, and values. The intent is to create dialogue that gives voice to rigorous thinking about possible meaning. Seminars are structured to take the student thought from the unclear to the clear, from the unreasoned to the reasoned . . . from the unexamined to the examined.'

(Lambright, 1995, p.30).

Socratic questioning is guided by a number of principles. One of the main principles of the method is being 'Inductive' (Mac Donald Ross, 1993). Induction occurs when specific cases are used to produce generalizations, as opposed to deductive reasoning where specific conclusions are drawn from generalizations. Another principle of the Socratic Method is how 'Language' is used to construct an argument. As mentioned, Socrates had some similarities to Sophists in that they both studied argument. But Sophists taught defendants how to represent themselves in the Greek legal system by manipulating the meanings of words to sway arguments in their favour (sophistry), and they provided this advice for financial remuneration. Socrates' purpose was quite different; he wished to use language to develop the validity of arguments to get at the truth, not just to win the argument. However, both the Sophists and Socrates did agree on one important issue and that was on the importance of language in thinking and rhetoric.

Another principle that Socrates adhered to in his teaching was to remain 'Open'. The effect of this principle was that all topics became open to revision. Socrates also believed in each person's entitlement to access philosophy and suggested it should be 'Applied' to everyday scenarios. This drive for application, combined with the

previous principle of openness, ensured his words were spread to a wide audience. Like the modern critical thinking philosophers such as Ennis and Paul, Socrates hoped his version of critical thinking could be useful for everyone in his or her work and life.

The works of Socrates, Plato and Aristotle went on to form the basis of the early Academies in the Middle Ages. Two Franciscan scholars from Merton College in Oxford University, John Duns Scotus (1270 - 1308) and William of Ockham (1280 – 1349) were typical of those in the early Academy who commented on a need for critical thought processes. Ockham's razor is still a widely used principle when applying critical thought - when deciding between two theories or explanations prioritize them by using the simpler one first. Examples of critical thought can also be seen in many of the classic works published since the dark ages. Thomas Aquinas (1225 -1274) in his seminal work, *Summa Theologica*, used a style of writing, which displayed critical reflection about his own work. He proposed the criticisms, which he thought others might have of his own theological and philosophical theories, and then tried to defend his ideas in the light of these imagined criticisms - a form of argument and counter argument. Galileo (1564-1562) could be considered the originator of scientific thinking/hypothesis testing and evidence-based reasoning, which is another skill linked with critical thought. Hegel (1770-1831) too demonstrated his belief in the importance of critical thinking and the dialectical processes advocated by Socrates (Frost, 1942). Francis Bacon (1551-1626) and René Descartes (1596 - 1650) continued to promote the principles of critical thinking. Bacon reintroduced the Greek notion of theory being tested by empirical observation (evidential support) and Descartes (1596-1650) suggested that thinking should be

doubted and questioned, again Socratic ideals (Wudka, 1998). Charles Darwin (1809-1882) applied a similar style of writing to Thomas Aquinas when he tried to refute his own criticisms in *The Origin of the Species* (1859). In general, critical thinking has strong roots in the western philosophical tradition.

The early part of the 20th century saw John Dewey (1859 - 1952) develop his theories of inquiry which are strongly anchored in the theoretical notions of critical thinking. Shermis (1999) suggests the current widely used psychological terms of ‘problem solving’, ‘higher level thought’ and ‘critical thinking’ are hybrid ideas stemming from John Dewey’s idea of ‘Reflective thought’ as discussed in his 1909 book *How We Think* (1993 revision). Dewey defined reflective thought as;

“Active, persistent, and careful consideration of any belief or supposed form of knowledge in the light of the grounds that support it and the further conclusion to which it tends” (Dewey, 1909, p.9).

This definition initially points to the substantial effort required for good thinking; it also emphasises the need to have good grounds or evidence to support a position - in the Socratic sense. Lastly, it highlights the importance of evaluating information or knowledge when drawing a conclusion. In short, this definition could well double as a contemporary definition for critical thinking.

Advocacy for critical thinking is not confined to Western traditions. Amartya Sen (2001), a Nobel Prize winning economist, argues that the Eastern tradition of critical thinking is as old as that of the West. Sen reports that the Indian emperor and scholar

Ashoka (273-236 BCE) was an advocate of critical thinking and tolerance. Tscherbatsky (1962) and Matilal (1990) also point out there has been a history of logical thinking in India for over two thousand years, which arose independently of Aristotelian philosophy. Tscherbatsky (1962) suggests the debates on the future of certain monasteries by the two legendary Buddhist logicians Dignaga (440 AD) and Dharmakirti (600-660 AD), exemplify this position.

Conversely, it has been suggested that there is a cultural barrier to critical thinking in the East (Atkinson 1997, Davidson 1995) because of the collectivist and hierarchical nature of many eastern societies where the teachings of the Chinese philosopher Confucius (551-479), has had a particular impact (e.g., China, Japan, Korea, Vietnam, Hong Kong and Taiwan). Confucianism is based on stability and authority which supports unequal relationships between people (Martinsons and Martinsons, 1996) and, it is argued, these unequal relationships have a negative effect on the development of critical thinking (Wollam, 1992). It is felt that strict parenting, rote learning and silent unquestioning classrooms advocated by Confucian philosophy, promote passive acceptance of knowledge and an unquestioning or uncritical attitude (Fielding, 1997).

Despite this, there are contemporary movements in the East which mirror critical thinking movements in the West and this may be due to the increasing democratization of many eastern countries. In Thailand, for example, legislation is being proposed that their education system should develop higher order thinking abilities to include critical thinking (Atagi, 2004). Recently, the Universiti Brunei Darussalam in South-East Asia advertised for researchers in an expanding critical

thinking department which focuses on some of these east- west issues on critical thinking. Furthermore, the Thai philosopher, Hongladarom (1998), agrees with western philosophers like Ennis (1996a) and Paul (1993) that there is a global requirement for critical thinking in modern society because of the increasingly diverse views to which people are exposed.

1.4 Critical Thinking Theory: Born in Philosophy - Adopted by Psychology

Philosophers in the United States have carried out most of the work over the last 40-50 years in the ‘critical thinking’ field. The work has grown at such an exponential rate that it has been referred to as the ‘critical thinking movement’ (Barnes, 1992). The philosopher, Robert Ennis, is considered to be the father of the current critical thinking movement in the US, which was sparked by his article ‘A concept of critical thinking’ in the *Harvard Educational Review* (1962). In this article, Ennis attempts to clarify the notion of critical thinking and he proposes a taxonomy of twelve critical thinking constructs. He also attempts to describe the relevance of each of these twelve constructs to teaching and assessment. Ennis has published many further articles on the nature and application of critical thinking (Ennis, 1962, 1985, 1987, 1990, 1992 and 1996a,). He has also been explicitly influential in its assessment (1993 and 1996b) and has developed two ability tests of critical thinking, namely, the Ennis-Weir Critical Thinking Essay Test (Ennis and Weir, 1985) and the Cornell Critical Thinking Tests (Ennis and Millman, 1985a and 1985b), (see Chapter 3 for greater detail on these tests and other critical thinking tests.)

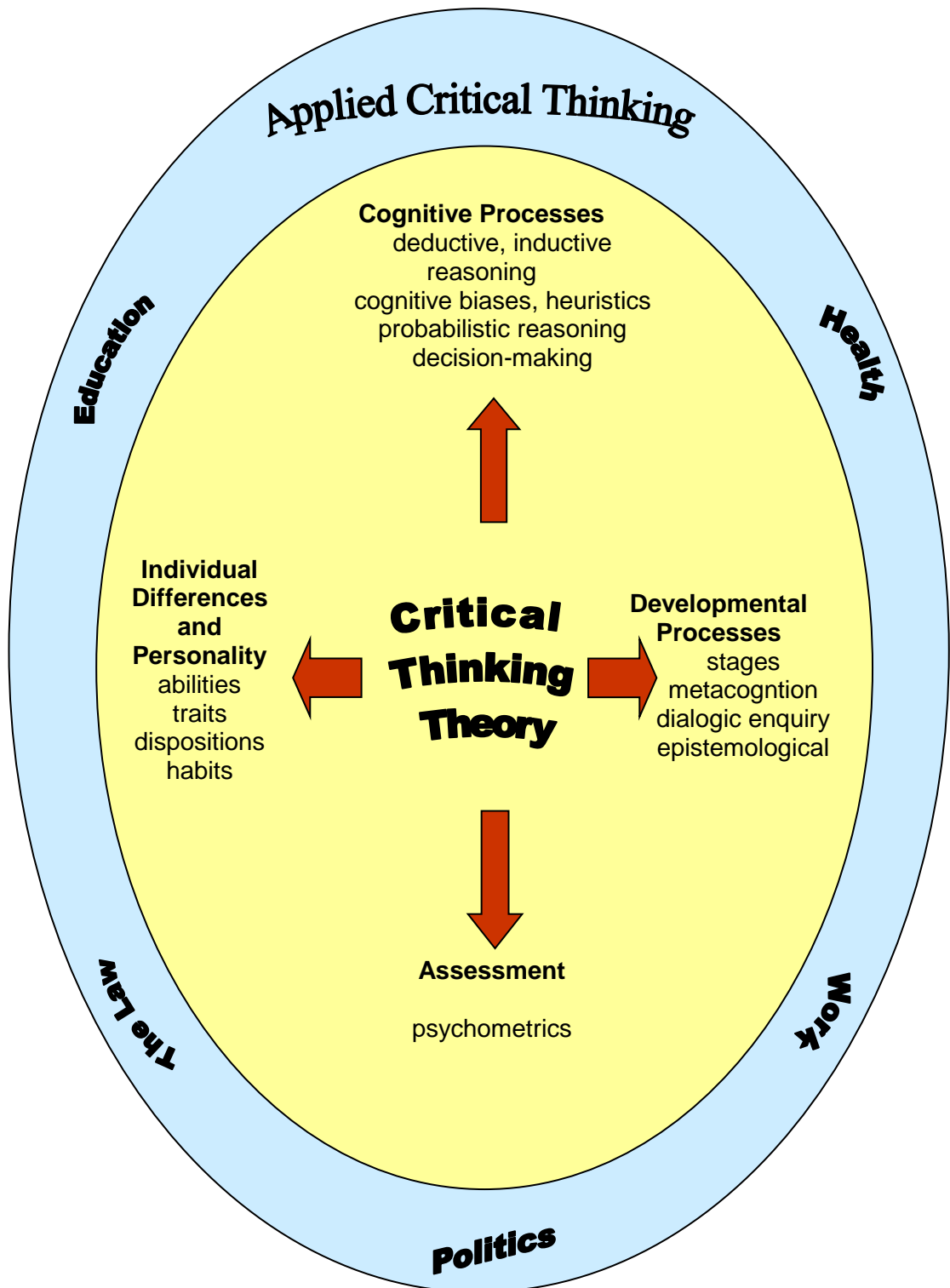
Currently, there are many other philosophers continuing to contribute to discussions on the nature and application of critical thinking skills, including Richard Paul

(2001), Bob Swartz (1994), Matthew Lipman (1987) in the US and Alec Fisher and Michael Scriven (1997) in the UK, as well as many others throughout the world. In essence, these philosophical discussions of critical thinking have continued in various guises for more than two millennia. An earlier interest in informal logic and the development of good arguments can be traced to two influential authors in this field, the English philosopher, Toulmin whose book *The Nature of Argument* (1963) discusses the use of argument in everyday life. Furthermore, the book, *Straight and Crooked Thinking*, by Thouless (first published 1930) describes the components of a good argument and also highlights the fallacies to look for in a poor argument.

More recently, psychologists have taken up these philosophical ideas and tried to add more empirical study to the arguments. Psychology as a discipline has both a conceptual and methodological toolbox that might prove useful for analysing critical thinking. As well, it has a history of systematic empirical investigation of constructs that are related, if not identical, to critical thinking. There is now considerable overlap between philosophical discussions of critical thinking and systematic empirical investigations by scholars from both disciplines. The purpose of the next part of this chapter is to impose some meaning and structure on these interconnections, to point out where the links are established and where there are gaps.

So where have the links with psychology been made? Figure 1.1 outlines where psychology has made important links with critical thinking theory (inner circle). The outer circle shows how wide-ranging the applications of critical thinking can be. The links in the inner circle of Figure 1.1 will be considered first.

Figure 1.1 Linkages between Critical Thinking Theory, Sub-disciplines of Psychology, and Critical Thinking Applications



Most definitions of critical thinking make references to the cognitive processes of reasoning in some form or another – deductive reasoning and inductive reasoning. So the first link to be explored is to cognitive psychology. If you open a cognitive psychology textbook, you are not likely to find a chapter entitled ‘critical thinking’. Instead, the relevant chapters will be entitled deduction, induction, evaluating syllogisms and arguments, mental models, cognitive biases and heuristics, probabilistic reasoning, and so on (Eysenck and Keane, 2000; Sternberg, 2003). There is a substantial amount of both theory and empirical work that is related to critical thinking. Halpern (1984, 1994, 1996 and 2002) was one of the first cognitive psychologists to systematically reconfigure current research in cognitive psychology and link it explicitly to the term critical thinking, to activities related to the critical thinking movement, and to point out the possible contributions that cognitive psychology could make to education for critical thinking.

Halpern’s contribution is discussed in more detail in Chapter 2. Cognitive researchers have also postulated on the neurobiology of thinking processes in the brain (Sala and Logie, 1998) and some more specific research has connected brain function with reasoning, particularly ‘hypothetico-deductive reasoning’ (Lawson, 2004).

Developmental psychologists have also theorized in the area of critical thinking. One of the major authors in this area is Deanna Kuhn. Kuhn (1998) proposes a model of the developmental stages of critical thinking in a Piagetan style, where she has highlighted qualitative changes in critical thinking ability and when these changes occur. She has studied the critical thinking development of all age groups from

young children to the elderly. Kuhn's important contribution to critical thinking research is discussed in more detail in Chapter 2. Fields of enquiry related to the development of critical thinking focus on the role of metacognition and language as a tool for mediating the development of critical thinking (dialogic thinking).

Psychometrics and Individual Differences are related psychological fields which further contribute to the field of critical thinking. Psychometricians develop psychometric tools, which can factor out and add statistical weight to theoretical components of critical thinking. They can also assess the reliability and validity of the measurement of these constructs. Furthermore, such tests can be used to evaluate the extent to which critical thinking has been developed or transferred after a critical thinking intervention. Facione (1991) has developed some of the few psychometric measures, which assess critical thinking and his tools will be discussed in greater detail in Chapter 4 and investigated in Chapters 6, 7 and 8.

1.5 Critical Thinking Applications: Education and Beyond...

It is not surprising that the importance of critical thinking and its development is widely recognised in education at all levels. Bloom's (1956) taxonomy of educational objectives reminds us that education can favour different forms of learning – knowledge, application, evaluation, comprehension, analysis and synthesis and that there is more to education than memorising knowledge. In the US during the 1980s, there was an explosion of interest in developing forms of higher-order thinking (including critical thinking) and a substantial research base and practice was established (e.g., Nickerson, Perkins and Smyth, 1985). In fact a whole issue of the journal *Teaching of Psychology* (Halpern and Numedal, 1995) focused solely on this

topic. Also, many international conferences on developing thinking, led by US researchers and scholars, have been running for the past twenty years (e.g., International Conference on Critical Thinking, organised by Richard Paul in California, has a nineteen year history and the 12th biannual International Conference on Thinking will be held in Melbourne in July 2005).

In the UK the interest in developing thinking skills in school learning became more evident during the late 1990s. In a report commissioned by the Department for Education and Skills in London (then called Department for Education and Employment), McGuinness (1999) reported and evaluated a range of approaches for developing pupils' thinking which were emerging in the UK; these included examples of the use of Lipman's *Philosophy in the Classroom* (1980) with its emphasis on Socratic enquiry. She concluded that "if we want students to become better thinkers, we must make explicit what we mean by these better forms of thinking and devise ways of educating directly for thinking". (McGuinness, 1999, p.5). Planned revisions to the Northern Ireland Curriculum will include an emphasis on both critical and creative thinking skills (CCEA, 2003).

In higher education, critical thinking has always been an important aim and a number of recent developments have made it more prominent. During 1990s each discipline in higher education was required to develop benchmark statements identifying learning outcomes for graduates in the discipline (HMSO, 1997). These statements include references to knowledge and understanding, subject-specific skills and generic skills. Critical thinking featured highly as a desired learning outcome across all disciplines and, traditionally, the highest levels of critical thinking are judged to

be associated with first-class degree performances (see for example, marking guidelines from the British Psychological Society, 2003). Many lecturers and researchers in higher education are actively looking for ways to develop and assess students' critical thinking skills and dispositions. Workshops on critical thinking are being organised by learning and teaching networks (e.g., Learning and Teaching Subject Network in Psychology in October, 2002 in Belfast); critical thinking is becoming a recognised topic for symposia at learning and teaching conferences (e.g. Psychology Learning and Teaching Conference, Strathclyde in April 2004) and publications reporting critical thinking interventions and assessment in UK contexts are appearing (Anderson and Soden, 2001; Ramiene, 2002).

These developments raise important theoretical questions about the nature of critical thinking. Do students have a general critical thinking ability or is it domain-specific? Are there benefits to be gained in designing 'separate' courses for developing critical thinking or is critical thinking inextricably linked with knowledge and understanding in a specific domain (as research on expertise would indicate). There have been extended philosophical debates on the question (McPeck, 1981 and 1990; Quinn, 1994) with little resolution. In spite of this, the development and evaluation of courses on critical thinking continues around the world (e.g., Fitzgerald, 2000; Tsui, 1999; Van Gelder, 2001). Also, what are the implications of these debates for the assessment of critical thinking? Is it worthwhile devising assessments/tests of general critical thinking abilities, or should we leave critical thinking to be detected as part of the normal assessment methods used in higher education (essays, exams, projects and so on)? What is the role of critical thinking assessment not only for graduate learning outcomes, but for selecting students as suitable for higher education?

For a long time, critical thinking ability has been the focus of selection procedures for high level employment (e.g., in the civil service, management). The first named test of critical thinking, Watson-Glaser Critical Thinking Test (first published 1964) is used extensively in the selection of management level employees (see section 3.3.2). Also, one of the major commercial test producers, Oxford Psychological Press, has recently issued a new battery of occupational selection instruments, called Critical Reasoning Skills Series (Oxford Psychological Press, 2000). In some rather specific situations, business is expressing an interest in methods of analysing and recording the process of reasoning and critical thought. For example, in Australia, van Gelder (2001) has developed an argument mapping software (Reason Able) that is being used to develop critical thinking processes in business executives. In the US, Facione (1991) has a range of services on critical thinking assessment available from his publishing house, Insight Assessment (2003).

Forms of critical thinking are being identified as important in a wide range of professional practice in the UK. For example, in medicine and education there is an increasing drive for evidence-based and evidence-informed practice. This requirement demands that medical and educational practitioners (e.g. GPs and teachers) have access to, and are able to make, 'critical' judgments about research evidence that is relevant to their area of practice. In medicine (and to a lesser extent in education), this has led to methods for critically reviewing research papers (systematic review) so that clinical guidelines for best practice can be identified (e.g., Cochrane Collaboration, 2004 and Campbell Collaboration 2004).

Also, research is beginning to identify the importance of critical thinking for legal reasoning. From a critical thinking perspective, Kuhn, Weinstock and Flaton (1994) found variation in the reasoning strategies adopted by jurors that resulted in extreme verdict choices with the potential to undermine public confidence in the outcomes of jury deliberation.

In conclusion, it appears that a focus on higher order thinking in general and critical thinking in particular, is enjoying a new wave of interest in education and professional practice in the UK. This resonates with developments around in the world, particularly in the US and Australia.

1.6 Conclusion

This chapter shows how the theories on argument and logic of Greek philosophers have links with contemporary theories of critical thinking. For example, the ideas about evidence in support of reasoning, argument and counter-argument, questioning and dialogue, clarity and openness are constructs that have survived and are evident in modern critical thinking discussions.

The recurring theme of this chapter is that psychology as a discipline can add additional theoretical perspectives and a more systematic and empirical method of inquiry, to the ideas and models of critical thinking proposed by philosophers and other thinkers. Furthermore, psychology's input from sub-disciplines such as cognition, development, individual difference and psychometrics have linked critical thinking analyses with other forms of higher-order thinking and enhanced critical thinking research.

More generally this chapter also reveals the diversity of both the theory and application of critical thinking and begins to show how important critical thinking is for living in the modern world.

Chapter 2

Critical Thinking: Definitions, Concepts, Frameworks and Models

2.1 Introduction

The last chapter presented a broad guide to critical thinking and pointed out the scope of the research activity on critical thinking and related questions. In this chapter the focus is more detailed. It attempts to critically evaluate the theories of major authors within the area of critical thinking. These authors range in discipline from educational philosophy to cognitive and developmental psychology and they have diverse perspectives on critical thinking as well as different ways of theorising. The educational philosophers' theorising is discursive, analytical and taxonomic, where expert judgement, identifying disputed areas, and logical argument are a means at arriving at theoretical consensus. In contrast, the psychologists are embedded in a more positivist and empirical tradition and tend to operationally define constructs and seek evidence to confirm or disconfirm hypotheses and theoretical positions. At times it can be difficult to reconcile the conclusions from these different modes of enquiry, even when they are referring to the same constructs. Nevertheless, despite the differences between them, the chapter attempts to show that there is substantial common theoretical ground between the different positions, as well as some differences.

Five theorists have been identified for analysis – Robert Ennis, Richard Paul, Alec Fisher, Diane Halpern and Deanna Kuhn. They are chosen because they represent major positions on critical thinking research and scholarship both in the US and the

UK. Ennis and Paul are major figures in the critical thinking movement in the US and Alec Fisher represents the same tradition in the UK. Halpern is a cognitive psychologist and her contribution comes from the extent to which she has related the general field of cognitive psychology to work on critical thinking rather than from her own primary empirical research. In contrast, Kuhn is a developmental psychologist who has conducted primary research on the development of children's and adults' reasoning and has begun to frame these results into a more general theory on the development of critical thinking and epistemological beliefs. As well as providing some detailed analysis of specific theories, the purpose is to draw out similarities and differences between them. Despite differences in terminology and the theoretical status of some of the terms, it is possible to identify recurring concepts and themes. At the end of the chapter, an attempt is made to pattern the most frequently occurring constructs (Figure 2.1) to get some overall impression of the meanings of critical thinking and how they are related to the constructs to be assessed in this thesis.

Before beginning on the detailed analysis of individual theories, the chapter outlines the contribution of a report, published in 1990, of systematic discussion using the Delphi method, on the nature of critical thinking which was chaired by Peter Facione (Facione, 1990). Because of the important distinction between critical thinking skills and critical thinking dispositions which emerged from that report, their conclusions will be used to frame much of the discussion in the chapter.

2.2 Critical Thinking: A Statement of Experts' Consensus for Purposes of Educational Assessment and Instruction: 'The Delphi Report' (Facione, 1990)

One of the most important developments in the literature on critical thinking in the past 15 years was the publication of an 'expert group' using the Delphi methodology (Gordon, 1994). Delphi is a technique used to gauge the degree of agreement/disagreement among experts on any topic where there are existing theoretical frameworks but some degree of confusion about terminology, meaning, emphasis and so on. Essentially, a first set of questions are identified and answers/views on the questions are elicited from a panel of experts. The answers are then pooled by the coordinator/chair of the discussion, areas of agreement/disagreement are identified and the cycle is repeated until maximum agreement is reached and the remaining areas of disagreement are outlined. An important feature of the technique is that consensus is reached by the experts offering an opinion and then leaving that opinion open to change in the light of other panelists' suggestions and argument. The results of the exercise are usually presented as the percentage of the panellists who agree with the final statements.

A Delphi exercise was initiated by Peter Facione in 1988 and continued over a period of 1 year 10 months. There were six cycles of consultation/feedback with the panel. The expert panel consisted of 46 critical thinking experts including Robert Ennis, and Richard Paul (whose ideas on critical thinking are reviewed later in this chapter); 52% of the panellists were philosophers, 22% were educationalists, 20% were in the social sciences and 6% were physical scientists (only 1 expert dissented from the final recommendations of the report). The report begins by stating that the 'heart of education lies...in the process of inquiry, learning and thinking rather than in the

accumulation of disjointed skills and senescent information', echoing the sentiments a number of educationalists (Ennis, 1996; de Bono 1974, 1994; Paul, 1993). The main purpose was to identify and clarify the elements of critical thinking that are appropriate to a student in third-level education (college level in US terminology) so that critical thinking in higher education could be taught and assessed more effectively. Several of the recommendations from the report focus on education and assessment.

The main conclusion from the report draws a distinction between critical thinking skills and critical thinking dispositions which is an assertion supported by many other researchers (Baron 1985; Colucciello, 1997; Dewey, 1930; Ennis, 1987, 1991 and 1996b; Facione and Facione, 1992; Facione, Facione and Giancarlo, 1998; Facione, Giancarlo, Facione and Gainen, 1995, Martin, 1992; Giancarlo and Facione, 2001; McPeck 1991; Norris, 1992; Norris and Ennis, 1989; Passmore, 1967; Paul, 1993; Perkins, Jay and Tishman; 1993, Resnick, 1987 and Siegal, 1988). With this degree of consensus, the distinction is widely acknowledged as theoretically sound and useful in educational settings. It also reflects the more general psychological distinction between intellectual abilities and personality traits/dispositions that pervades the research literature on individual differences. However, the exact relationship between these two critical thinking components is far from settled and some of the debates will be discussed later on in this chapter.

The report defines a critical thinking skill as follows; 'A CT skill, like any skill, is the ability to engage in an activity, process or procedure. In general, having a skill includes being able to do the right thing at the right time. So, being skilled at CT

involves knowing, perhaps implicitly or without the ability to articulate this knowledge, both a set of procedures and when to apply those procedures. (Facione, 1990, p.31). Table 2.1 below taken from the Delphi Report (Facione, 1990, p.15) identified the main elements of critical thinking that were agreed upon by the expert group. Ninety-five percent agreed that *analysis*, *evaluation* and *inference* were central to critical thinking, while 87% agreed on *interpretation*, *explanation* and *self-regulation*. Each of the main skills in the agreed taxonomy were linked to sub-skills. The experts also agreed that all of these skills do not need to be present in order to demonstrate critical thinking but that they work in a dynamic and integrated way to produce critical thought.

Table 2.1: Consensus list of critical thinking cognitive skills and sub skills

Interpretation	Categorization
	Decoding Significance
	Clarifying Meaning
Analysis	Examining Ideas
	Identifying Arguments
	Analyzing Arguments
Evaluation	Assessing Claims
	Assessing Arguments
Inference	Querying Evidence
	Conjecturing Alternatives
	Drawing Conclusions
Explanation	Stating Results
	Justifying Procedures
	Presenting Arguments
Self-regulation	Self-examination
	Self-correction

This first main skill of *interpretation* consists of the sub-skills categorization, decoding significance, and clarifying meaning. The report suggests *interpretation* is ‘To comprehend and express the meaning or significance of a wide variety of

experiences, situations, data, events, judgements, conventions, beliefs, rules, procedures or criteria.’ (Facione, 1990, p.16)

The *Analysis* sub-skills are examining ideas, identifying arguments, analyzing arguments. The definition given for analysis is ‘to identify the intended and actual inferential relationships among statements, questions, concepts, descriptions or other forms of representation intended to express beliefs, judgments, experiences, reasons, information, or opinions.’ (Facione, 1990, p. 17)

Evaluation is composed of assessing claims and assessing arguments and is defined as the ability ‘to assess the credibility of statements or other representations which are accounts or descriptions of a person’s perception, experience, situation, judgement, belief, or opinion; and to assess the logical strength of the actual or intend inferential relationships among statements, descriptions, questions or other forms of representation.’ (Facione, 1990, p.18)

The *Inference* sub-skills are querying evidence, conjecturing alternatives and drawing conclusions. Inference is defined as the ability ‘to identify and secure elements needed to draw reasonable conclusions; to form conjectures and hypotheses; to consider relevant information and to educe the consequences flowing from the data, statements, principles, evidence, judgements, beliefs, opinions, concepts, descriptions, questions, or other forms of representation.’ (Facione, 1990, p.19)

The penultimate skill *Explanation* consisted of the sub-skills stating results, justifying procedures and presenting arguments. The report outlines this skill as the ability ‘to state the results of one’s reasoning; to justify that reasoning in terms of the evidential, conceptual, methodological, criteriological and contextual considerations upon which one’s arguments were based; and to present one’s reasoning in the form of cogent arguments. (Facione, 1990, p.18)

The last skill of *self-regulation* is composed of self-examination and self-correction and is very similar to metacognition in its content. It is defined in the Delphi report as ‘Self-consciously to monitor one’s cognitive activities, the elements used in those activities and the results deduced, particularly by applying skills in analysis and evaluation to one’s own or another person’s thinking on a matter of deep personal concern’. (Facione, 1990, p.22)

The second focus of the group’s discussion was on critical thinking dispositions. A critical disposition was defined as ‘the personal traits, habits of mind, attitudes or affective dispositions which seem to characterize good critical thinkers’ (Facione, 1990, p. 23). A crucial point made in this section is that, if a critical thinking skill is carried out appropriately, a correlated critical thinking disposition will be present. Also provided in this section is a description of the dispositions of a good critical thinker. Furthermore the Delphi report (Facione, 1990, p.28) identifies a list of the nineteen affective dispositions of critical thinking (see Table 2.2). This list is broken down into ‘approaches to life and living in general’ and ‘approaches to specific issues, questions or problems’. 61% of the experts agreed that the listed dispositions

were part of the conceptualization of critical thinking and 83% agreed that a good critical thinker could be ‘characterised as exhibiting these dispositions’.

Table 2.1: List of the ‘Affective Dispositions of Critical Thinking’ as outlined in the Delphi Report

Approaches to life and living in General
Inquisitiveness with regard to a wide range of issues
Concern to become and remain generally well-informed
Alertness to opportunities to use CT
Trust in the process of reasoned inquiry
Self-confidence in one’s own ability to reason
Open-mindedness regarding divergent world views
Flexibility in considering alternatives and opinions
Understanding of the opinions of other people
Fair-mindedness in appraising reasoning
Honesty in facing one’s own biases, prejudices, stereotypes, egocentric or sociocentric tendencies
Prudence in suspending, making or altering judgements
Willingness to reconsider and revise views where honest reflection suggests that change is warranted.
Approaches to Specific Issues, Questions or Problems
Clarity in stating the question or concern
Orderliness in working with complexity
Diligence in seeking relevant information
Reasonableness in selecting and applying criteria
Care in focusing attention on the concern at hand
Persistence though difficulties are encountered
Precision to the degree permitted by subject and circumstances

The table above shows the Delphi agreed component dispositions of the ideal critical thinker. However, they do not suggest that third level students necessarily have the level of personal development to match this ideal; but they do suggest that helping the student strive towards this ideal is the important goal. Some of the experts expressed the view that over-emphasis on the dispositions could lead to problems while others argued that neglect of this dimension could result in producing ‘close-minded, intellectually inflexible and dogmatic’ dispositions in the students (Facione, 1990, p.29).

Like many of the frameworks discussed in the remainder of Chapter 2 the two taxonomies described in Table 2.1 and 2.2 of the Delphi report have many linked parts and the skills and sub-skills are not mutually exclusive. Furthermore, many of the critical thinking skills and dispositions themes described in the Delphi report are similar to those seen in the models described in the remainder of the chapter. It is for this reason that the Delphi conceptualization is compared (section 2.9) with the model summary that occurs at the end of this chapter.

The Delphi Report also discusses critical thinking issues and goals relevant to education. How are critical thinking skills enhanced in an educative setting? How are critical thinking skills assessed? The 'Delphi Report' provides recommendations on how to teach critical thinking skills. There is recommendation to develop critical thinking as a transferable entity as this set of skills and dispositions pervade many aspects of personal and civic life. There is also a request to make the secondary education curriculum preparatory for students entering into third level domain, which is explicitly motivated to develop critical thinking. Another reason to include critical thinking in a secondary curriculum is that a large percentage of students will not enter tertiary education.

The main recommendation on critical thinking assessment is that it should be reliable and valid. The report also suggests that content and construct validity are crucial. Another explicit statement that the panellists agreed on is that the assessment should be fair. In other words the assessment techniques should not discriminate on variables such as gender or culture. The report states that if critical thinking assessment is carried out appropriately it can be a useful tool in directing educational

policy and curriculum formation. The final recommendations in the report concern the methods and goals of a critical thinking instructor. The experts concur that a critical thinking teacher should promote and utilise all the aspects of the critical thinking ethos instead of relying on outdated teaching methods such as rote learning.

Chapter 3 details two critical thinking tests that Facione developed using the information obtained from the Delphi group. The first test is a test of critical thinking skills (California Critical Thinking Skills Test, CCTST) and the second is a measure of critical thinking dispositions (California Critical Thinking Dispositions Inventory, CCTDI). These tests are independent of the Delphi Report and the Delphi groups' endorsements.

2.3 Robert Ennis

2.3.1 Definitions of critical thinking: Robert Ennis is a US philosopher and has been writing about critical thinking for over 40 years. He was a Professor of Philosophy in the University of Illinois for 25 years where he is currently a Professor Emeritus. Ennis sees critical thinking as essential to the proper functioning of a democratic society (Ennis, 1996).

“Critical thinking is the correct assessment of statements” (Ennis, 1962)

“Critical thinking is reasonable and reflective thinking that is focused on deciding what to believe or do.” (Ennis, 1981)

Ennis’ critical thinking definitions and frameworks have changed over time. Two versions of his definitions are provided above. A possible reason for his definition persisting despite the changes in his critical thinking frameworks may be that it is the most widely recognized and used definition of critical thinking. Therefore he may be reluctant to change such a popular definition.

2.3.2 Ennis’ Theories on Critical Thinking: Ennis explicitly discusses critical thinking dispositions and abilities. He also makes reference to a critical thinking checklist called FRISCO. The following paragraphs provide a brief description of Ennis’s theories on these three critical thinking components.

Ennis’ thoughts on critical thinking dispositions have developed through the years and he now sees critical thinking dispositions as important as critical thinking

abilities (1996b). Ennis suggests there are two essential dispositions in order to become a 'good' critical thinker and one desirable component. He further points out that these components are not mutually exclusive and overlap. However he has separated them for simplicity of explanation (Ennis 1996b).

The first of Ennis' two crucial components is entitled '*get it right*'; in other words, critical thinkers have a disposition to care that their beliefs are true to the best of their knowledge and can be justified with evidence. In essence, this disposition governs scientific thinking or logical reasoning. The second disposition of a critical thinker outlined by Ennis focuses on a person's responsibility to '*represent a position honestly and clearly (theirs as well as others)*'. In essence, this disposition relates to the appropriate use of language in arguments, to discern truth rather than just to win the argument, which can be traced back to Socratic influences (see section 1.3). The third disposition Ennis (1996b) describes as '*correlative and constitutive*' or desirable characteristics of a critical thinker. He names this disposition as '*care about the dignity and worth of every person*'. It includes the disposition to actively seek and care about other people's views, to avoid intimating and confusing others (by being 'too clever') and by being concerned with other people's welfare and avoid hurting their feelings. Ennis (1996b) concedes that this latter disposition is not required to produce critical thought. However, he claims that critical thinking can be devalued or rendered valueless if the impact of critical thought on others is not considered. In summary, if the first two dispositions are concerned with scientific and philosophical matters, the third disposition points to the human and interpersonal consequences of critical thinking. In addition, Ennis suggests his dispositions reflect values associated with the disciplines of science, philosophy and humanities.

Ennis also describes a taxonomy of critical thinking abilities similar in structure to his list of dispositions. His first four stated ability components of the ideal critical thinker namely, 1. *Clarify*, 2. *Judge the basis for a decision*, 3. *Infer*, and 4. *Integrate abilities*, have remained relatively constant throughout the years, although there has been some change in their content. In the most recent version of Ennis' (1996b) critical thinking skills framework, '*use auxiliary critical thinking abilities*', has appeared as a further category of abilities. The following paragraph provides a brief description of each of these critical thinking abilities.

The first main ability, *clarify*, is broken down into a number of constituent abilities. They include the ability to identify the focal issue, analyse arguments, ask and answer questions appropriately and, lastly, form and judge definitions. Overall, Ennis (1987 and 1996a) states this is an advanced skill and he discusses a lot of detail on forms of definitions and, like Socrates, he is adamant that defining terms is a crucial skill. The next group of abilities come under the heading of basis, i.e., what is the basis for an argument. Ennis argues that ideal critical thinkers should be able to judge the reliability and validity of a source and secondly that s/he should be able to make their own credible observations. Ennis contends judging source credibility is a skill that can be developed well by educational programmes from Early Years right through to Higher Education. This ability is becoming increasingly crucial with the multitude of non-reviewed content available to a wide range of people on the internet, television and newspapers. *Inference*, Ennis' third higher order group of critical thinking skills, has four sub components. They are; 'identify unstated assumptions', 'deduce and judge deductions', 'induce and judge inductions to generalization and explanatory conclusions' and 'make and judge value judgements'.

Again he sees much overlap between these differing types of inference. The fourth division of abilities Ennis suggests is '*making suppositions and integrate abilities*'. Ennis defines suppositional thinking as the ability to think about a reason's quality without letting doubt or disagreement interfere in the process. He further suggests the ideal critical thinker should have the ability to integrate all the above critical thinking dispositions and abilities in order to defend an opinion or decision. The fifth and last of Ennis' ability headings is '*Use of auxiliary critical thinking abilities*'. The first sub-ability in this group is '*proceed in an orderly manner appropriate to the situation*'. He gives examples of this ability as a person who monitors their thinking and problem solving almost in a metacognitive sense. This sub-ability also includes the systematic process of using a critical thinking/reasoning checklist. Ennis outlines such a checklist called FRISCO, which is described in the next section. The second sub-ability here is being aware of others, feelings, abilities as critical thinkers and level of knowledge; so as not to intimidate or hurt them. This sub-ability has obvious ties to Ennis' third convergent critical thinking disposition (described above). The last ability in this group is the use of rhetorical techniques during discussion or presentation.

There have been a number attempts to summarise the components of critical thinking. The summary taxonomy Robert Ennis puts forward is called FRISCO checklist and it is described in detail (1996a). He suggests that his checklist is easy to remember and can be employed as a useful tool for encouraging critical thinking. Many others have recommended metacognitive style checklists for their potential in developing critical thinking particularly in younger children, Fisher (1990), Fisher and Scriven (1997), Halpern (1998), Lrynock and Robb (1999), Wilks (1995). The

FRISCO acronym superimposes itself onto Ennis' taxonomies but not precisely. The letters stand for Focus, Reasons, Inference, Situation, Clarity and Overview and the following paragraph gives a brief synopsis of each.

Focus directs the person focusing on the problem or situation. *Reasons* are akin to the ability of basis, i.e., support for one's inferences. Likewise *inference* is directly taken from the taxonomy of critical thinking abilities, i.e., deduction, induction etc. *Situation* promotes Ennis' notion of the need for critical thinking to be used for the purpose of good or, at least, not to harm anyone. *Clarity* too maps onto his taxonomy ability heading of the same title. *Overview* has some of the concepts of his fourth ability group 'integrate'. Here he encourages the critical thinker to tie all the other components and see if the belief or action is still justified. This is also a point of 'metacognitive' reflection for the thinker where they obtain awareness of what has been learned from the process.

2.3.3 Evaluating Ennis: Ennis refers to three general critical thinking components, i.e, skills, dispositions and metacognitive components. His three dispositions are 'to get it right', 'represent a position honestly and clearly' and 'care about the dignity and worth of every person'. His suggested critical thinking skills are Clarify, Judge the basis for a decision, Infer, Integrate abilities and use auxiliary critical thinking abilities. Ennis' metacognitive components concern awareness of what has been learnt while reflecting on both one's critical thinking skills and dispositions.

Ennis suggests that this is not an exhaustive list of critical thinking components. He also argues that there may be more critical thinking abilities that are domain specific

to certain knowledge disciplines. Furthermore, he states there is much overlap in his lists of dispositions and abilities and he concedes that his taxonomy is ‘not an elegant list; this is a practical list’ (Ennis 1987 p.18).

Overall, Ennis’ ability lists are complex due to their detail, mutual non-exclusivity and mutation, as he has moved his focus from critical thinking skills to critical thinking dispositions. Furthermore, his theories have little or no empirical support. However, like Bloom’s (1956) taxonomy he has provided an influential list of critical thinking abilities and dispositions, which has sparked much debate. In short, more than any other author, Robert Ennis’ ideas have promoted the area of critical thinking.

2.4 Richard Paul

2.4.1 Richard Paul’s Definition of Critical Thinking: Like Robert Ennis, Richard Paul is a philosopher and is recognized as one of the leaders of the critical thinking movement in the US. He is currently the director of research at the *Center for Critical Thinking and Moral Critique*. He has lectured extensively throughout the world on critical thinking and he has run seminars on critical thinking for many professional groups including teachers, business executives, doctors and theologians. He has also hosted the *International Conference on Critical Thinking at Sonoma State University* for the last nineteen years. Paul is passionate in his belief that educational reform is the solution to social and economic problems.

Definition 1

‘Critical thinking is disciplined self-directed thinking which exemplifies the perfections of thinking appropriate to a particular mode or domain of thinking..... If the thinking is disciplined to serve the interests of a particular individual or group, to the exclusion of other relevant persons or groups, I call it sophistic or weak sense critical thinking. If the thinking is disciplined to take into account the interests of diverse people or groups, I call it fair-minded or strong sense critical thinking..... As we come to think critically in the strong sense we develop special traits of mind: intellectual, humility, intelligence courage, ‘intellectual perseverance, intellectual integrity and confidence in reason.’

Richard Paul (1993)

This definition of critical thinking by Richard Paul’s (1993) is not as concise as Robert Ennis’ definition. However it arguably includes an extra element that Ennis’ definition does not include; yet it is a concept that Ennis would advocate. The extra element in the Paul definition is his advocacy of the proper use of critical thinking as a way of tackling society’s problems. An interesting aside is Paul’s use of the term sophistic, which refers back to the Greek origins of critical thinking type discussion as mentioned in Section 1.2. Paul refers to his weak sense (or low level) critical thinking as Sophistic. Overall, this definition of Paul’s concentrates on the ethos of critical thinking and the potential benefits to society more than what the cognitive components of critical thinking are. The following definition by Scriven and Paul (NCECT, 2004) discusses more precisely what Paul believes the structure of critical thinking to be.

Definition 2

Critical thinking is the intellectually disciplined process of actively and skilfully conceptualising, applying, analysing, synthesizing, and/or evaluating information gathered from, or generated by, observation, experience, reflection, reasoning, or communication, as a guide to belief and action.

Michael Scriven and Richard Paul (NCECT, 2004)

This definition reads like a taxonomy of critical thinking skills and is very unlike the previous definition. Many of the skills in this latter definition are mentioned by Ennis and the other authors evaluated in the remainder of this chapter.

2.4.3 Richard Paul's Models and Frameworks of Critical Thinking: Paul, like Ennis, has developed the structure and content of his theories of critical thinking over years. He has published two major theories of critical thinking. The first theory is a model called, *Strategy List for Redesigning Lessons* and his second theory is a framework, which describes, *what is essential to critical thinking and what it is to BE a critical thinker*.

His initial model *Strategy List for Redesigning Lessons* has 35 strategies, which are broken down into three sections which are, Cognitive Strategies - Macro Abilities, Cognitive Strategies - Micro Abilities and Affective Strategies. It is suggested the content of the two cognitive strategies groups are similar to critical thinking abilities as they feature skills such as analysing, questioning, reasoning and inference, while the affective strategies are similar to critical thinking dispositions, as they feature attitudes like fair-mindedness, intellectual humility and intellectual perseverance.

Paul's framework has four parts, with the first three described, as essential to critical thinking; they are Elements of Reasoning, Standards of Critical Thinking and Intellectual Abilities. The fourth group are called Intellectual Traits and are described as essential to BE a critical thinker. This framework, like his original model, can be divided into abilities and dispositions. The first three components, i.e., the essential to critical thinking components, could be described as his taxonomy of critical thinking abilities and the last category of intellectual traits which he also calls 'Traits of the disciplined mind' could be considered as his descriptions of critical thinking dispositions.

Paul and Elder (2001) suggest that the 'elements of reasoning' group contains the fundamental components for human reasoning. Paul sees the next group 'standards in critical thinking' as those components of human thought that are not commonly present. However mastery in these components produces quality critical thought. In essence the standards list contains skills like the elements of reasoning list, however they are considered to be higher order. Paul postulates that an intellectual ability is composed of three elements. The first is a *process* and he gives the example to drive. The second is an *object* and the example is a truck and the third is a *standard* in this example safely. So the intellectual ability example is to drive a truck safely. Intellectual traits are the dimension, which Paul suggests produce the *strong sense critical thinker* or *essential to what it is to BE a Critical Thinker*. He suggests that these are moral or affective components of critical thinking. Paul enhances his framework by placing them in two different contexts. The first is a six stage developmental process of critical thought. He suggests that a person can move from

the lowest stage of an unreflective thinker through to the final stage of a master thinker. He also proposes the notions of strong and weak critical thinkers, that is, people can use critical thinking for either altruistic reasons (strong) or personal benefit (weak).

2.4.4 Evaluating Paul: Like Ennis Paul, provides a model featuring both critical thinking skills (abilities essential to critical thinking) and dispositions (attitudes essential to BE a critical thinker). Paul has additional ideas based on a developmental scheme of critical thinking.

In general Paul's ideas are very complex. He has produced a number of theories on critical thinking that are very precise in their detail, yet have no empirical grounds. For example, his developmental scheme has no research support unlike Deanna Kuhn's developmental scheme, which is reviewed in section 2.7.2. The result of this complex web of theories, models and frameworks is that they are difficult to reconcile and thus working with them is a difficult exercise.

2.5 Alec Fisher

2.5.1 Alec Fisher's Definition of Critical Thinking: Alec Fisher is an English academic who is currently the Director of the Centre for research in Critical Thinking at the University of East Anglia, Norwich, UK. He has been partnered in most of his work by Michael Scriven who started his academic career in Australia and is now Professor of Psychology at Claremont Graduate School in California, USA. They have both worked with Richard Paul on critical thinking theory. Fisher and Scriven have also published a book together called *Critical Thinking: Its*

Definition and Assessment (1997). Alec Fisher also has published an undergraduate text-book on critical thinking *Critical Thinking: An Introduction* (2001). This is a preparatory text on critical thinking partly based on the original book by both authors. Although both authors have a background in mathematics, logic and psychology a lot of their work on critical thinking is philosophical in its nature and they have theorized on the nature of critical thinking in a manner similar to both Ennis and Paul. These two authors form some of the strongest links between the US and UK on Critical Thinking research.

“Critical thinking is skilled, active interpretation and evaluation of observations, communications, information, and argumentation.”

Fisher and Scriven (1997).

Fisher and Scriven describe this definition in extensive detail in their 1997 publication. They break it down and explain each term with examples. They suggest that evaluation is such an important part of this definition that they refer to it as the ‘*evaluative definition*.’ Fisher and Scriven also base their model of critical thinking on this definition and this model is described in the next section.

2.5.3 Fisher’s Model of Critical Thinking Competencies: In the 1997 text Fisher states that his framework of critical thinking is similar to the taxonomies of others like Ennis and Paul. However, the taxonomy is specifically derived from their ‘*evaluative definition*’. As a result he proposes a more focused model of critical thinking based on the central skill of evaluation. He suggests that, although these other taxonomies detail some of the essential aspects of reasoning, reading and

writing they do go beyond what should be seen as essential critical thinking components. He believes this model of critical thinking competencies describes what is at the 'cutting edge' of thought. Interestingly, the definition and model of critical thinking competencies does not include critical thinking dispositions. The reason he gives for this omission is that he is specifically describing the qualities of critical thinking and not the attitudes of the critical thinker. In other words he does not feel that the definition of critical thinking should include an affective or dispositional dimension and should remain simply as a list of cognitive skills. He suggests that critical thinking is best evidenced by behaviour rather than self-report. Fisher does not entirely disregard critical thinking dispositions as he does state that the teaching of critical thinking should include the development of the critical thinking attitudes.

Fisher has unpacked his 'evaluative definition' of critical thinking into a model which comprises four groups of competencies. They are *critical interpretation*, *critical communication*, *critical knowledge* and *critical technique*. *Critical interpretation* is apparent when the thinker actively scrutinises information whether the information is read, heard or observed. Fisher suggests the next category *Critical Communication* compliments the previous category interpretation as communication is the output where interpretation is the input. The communication group of competencies is broken down to include the person's ability to write, speak and present to audiences in a clear, comprehensive, concise, suitable, original and powerful way. *Critical Knowledge* describes awareness of critical thinking concepts and scenarios. He divides this section into three sub categories called the language used in critical thinking, basic vocabulary used in logic and knowledge of 'the missing in action' areas. The last ability called knowledge about the 'missing in

action' areas refers to thinking about issues that are not currently covered in educational programmes. Fisher explains that these topics do not make it into the educational system because they are so current and fiery that there is not time to make room for them in the congested, slowly changing curriculum. He further suggests this is unfortunate as these hotly debated topics are conducive to promoting this useful critical thinking competency. Fisher suggests that the competency group *critical technique* is composed of the 'skills' most associated with the critical thinking models. This group includes, interpretation of context, clarification of meaning, analysis of arguments and synthesis of considerations.

2.5.4 Evaluating Fisher: There is a strong sense that Fisher's model of critical thinking is based around the principles of informal logic. This is typified by his numerous referrals to the language and concepts of logical discussion. Fisher's model has a number of similarities and differences to the frameworks of Ennis and Paul. Many of the abilities identified by the two American philosophers mirror those mentioned in Fisher's model. For example, Fisher describes *Clarification of meaning* which overlaps with one of Ennis' skills i.e. *clarify*. Fisher also highlights the importance of integrating the various skills to use them in tandem. This hybrid use of skills is also advocated by Ennis and named a skill of '*making suppositions and integrate abilities*'. However, Fisher's model differs in one major respect from the others in that he excludes critical thinking dispositions from his framework.

Overall Fisher's model is the most focused that is examined in the thesis. It is based on the 'Evaluative definition' and thus has evaluation as its primary concern. However, this model excludes a number of critical thinking components put forward

by a number of influential authors. Furthermore, Fisher excludes them without the support of empirical evidence.

2.6 Diane Halpern

2.6.1 Diane Halpern's Definition of Critical Thinking: Diane Halpern is the Director of the Berger Institute for Work, Family, and Children and Professor of Psychology at Claremont McKenna College in Los Angeles. Halpern's seminal work 'Thought and Knowledge an Introduction to Critical Thinking' (first published 1984) is now in its 4th edition (2002). Halpern attended a US government-funded workshop in 1992 which was held in response to a statement by the National Education Goals Panel (1991) 'The proportion of college graduates who demonstrate an advanced ability to think critically, communicate effectively, and solve problems will increase substantially.' During that workshop Halpern suggested categories of thinking skills that now form the basis of her framework of critical thinking, which is described in more detail in the next section. Halpern also stated during that workshop that these skills are 'needed to compete in a global economy and in the exercise of citizenship.' In this statement Halpern shows the common position of Ennis, Paul etc. of the need for critical thinking in a proper functioning democratic society. However, Halpern differs from these philosophical writers in that she comes from the perspective of a cognitive scientist and her framework describes typical cognitive domains of research like reasoning and problem solving.

'Critical thinking is the use of those cognitive skills or strategies that increase the probability of a desirable outcome. It is used to describe thinking that is purposeful, reasoned, and goal directed – the kind of thinking involved in solving problems,

formulating inferences, calculating likelihood, and making decisions when the thinker is using skills that are thoughtful and effective for the particular context and type of thinking task.'

Halpern (1996, p.5).

Halpern describes her definition of critical thinking as a working definition, so that it can be applied to many real-life situations, which require critical thought. It is broad in its scope of cognitive domains but like Fisher she does not explicitly include critical thinking dispositions in the definition.

2.6.2 Halpern's Framework of Critical Thinking: The following description of Halpern's critical thinking framework is taken from two sources, namely her 1996 text and her input in a 1994 book which aimed to set standards in the assessment of college student learning (Greenwood, 1994). Her input in the 1994 book featured 'A taxonomy of critical thinking skills' (1994, p.31) which she prepared with the goal of producing a template for the national assessment of critical thinking skills in the adult population. Halpern's list of critical thinking abilities is very comprehensive consisting of at least 63 skills listed under 9 categories. Halpern names each of these skills, gives a description and an example of their usage. Although the following only provides a brief description of the 9 categories it is hoped that the reader will be able to position Halpern's work with the other authors described in this chapter.

The first of the nine categories in Halpern's taxonomy is called *Memory Skills*. This group is described as the skills required when learning, retrieving and retaining information. *Deductive reasoning skills* form the next group outlined by Halpern.

This category in her framework consists of the skills that are used for deducing conclusions from a premise. *Argument analysis skills* are skills that are needed to judge strength of reasons, counter-reasons, conclusions and overall arguments. Examples of skills in this category include, judging credibility of sources and understanding the difference between opinion, reasoned judgment and fact. The fifth group is named *skills in thinking as hypothesis testing*, which is described as formulating hypothesis or beliefs, collecting information in the form of observations, and using that information to confirm or disconfirm those hypotheses. The next category is *likelihood and uncertainty critical thinking skills*, which Halpern describes as using objective and subjective probability estimates in an appropriate manner. In other words, the thinker is aware of the benefits and limitations of statistical analysis. *Decision-making skills* are described by Halpern as the generation, selection and judgment of alternatives. *Problem-solving skills* are the penultimate group that Halpern proposes. She suggests this is the ability to focus on a problem, suggest goals and solutions to that problem. The ninth and last group Halpern outlines in her framework is '*skills for creative thinking*.' This group has skill elements such as brainstorming and visualization from differing perspectives.

Lastly, there is one striking omission in Halpern's framework, i.e., critical thinking dispositions. Halpern's discussion of critical thinking dispositions is not comprehensive. However, she does put forward a list of attitudes which she feels teachers should try and instil in their students. They include 'willingness to plan, flexibility (open-mindedness), persistence, willingness to self-correct, being mindful and consensus seeking'. Halpern does provide greater detail about a metacognitive aspect to critical thinking. Her metacognitive framework includes four questions.

1. What is the goal?
2. What is known?
3. Which skills will get you to your goal?
4. Have you reached your goal?

2.6.4 Evaluating Halpern: As stated, this framework has a strong cognitive influence, although Halpern has not carried out empirical research on critical thinking she does draw heavily on cognitive research to produce her theories on the subject. Skills like problem solving, decision making and memory should be particularly familiar to cognitive psychologists. Furthermore this is the first time these familiar cognitive skills has been included in the critical thinking frameworks reviewed in this chapter. Fisher and Scriven would generally argue against the inclusion of such a wide range of cognitive skills for a number of reasons. Fisher and Scriven (1997) state that critical thinking is closely related to decision making but they suggest that decision making is not actually a critical thinking skill. Good decision making is reliant on critical thinking but a person can still make decisions without using critical thinking. Fisher and Scriven (1997) would again argue against the inclusion of problem solving. They suggest in Mathematics, for example one often has to solve numeric problems yet it need not be described as critical thinking. Furthermore, they state that animals like dogs often solve problems but it would not be accurate to say they are doing critical thinking. However, others have suggested that problem solving should be included under critical thinking (Kurfis, 1988). More generally, Fisher and Scriven identify the problem of popularity as a reason for the inclusion of a wide range of cognitive skills within the critical thinking framework.

In order to gain recognition for their research, writers may be tempted to specify any higher-order thinking as critical thinking.

Halpern's model is not exclusively influenced by cognitive psychology, as she draws on the ideas of philosophers like Ennis, Paul and de Bono. As a result, skills included in her framework like argument analysis, thought and language skills and creative thinking have some philosophical notions. For example Halpern's thought and language skills feature many of the philosophical characteristics of Ennis' clarify ability. A further characteristic in Halpern's framework similar to the models of Ennis and Paul is that she states the categories are not mutually exclusive and often work in a dynamic way to produce critical thought. Halpern also suggests a general guiding checklist for critical thinking like Robert Ennis' FRISCO system. Halpern calls this a metacognitive framework. However, unlike Ennis and Paul, Halpern includes creative thinking theory in her model much of which has parallels to the work of de Bono. Fisher and Scriven would agree with Halpern on the inclusion of this construct as they state it is impossible to be a critical thinker without being able to create different perspectives and viewpoints on issues. However they warn that this type of creativity is not the same as artistic or writing ability which they suggest is not required to think critically.

In summary, Halpern's model can be loosely divided into three components, i.e., cognitive skills, traditional critical thinking skills (mirroring many of the constructs proposed by other major authors in the area) and a metacognitive dimension. Overall Halpern believes that critical thinking is more necessary now than ever before because of the vast role of the media (newspapers, advertising etc) and politics in our lives. Halpern's book (1996) offers many examples of everyday usage of critical thinking and the accompanying work book (2002) has many exercises which provide

scenarios in which critical thinking is required. She also believes that a nation can increase its wealth if critical thinking is developed.

2.7 Deanna Kuhn

2.7.1 Kuhn's Definition of Critical Thinking: Deanna Kuhn is Professor of psychology and education in the University of California at Berkeley. Her teaching covers critical thinking, cognitive development, research methods in developmental psychology and applied educational research. Her many publications areas include; scientific thinking (Kuhn, 1989; Kuhn, Amsel, and O'Loughlin, 1988; Kuhn and Pearsall, 2000; Kuhn, Schauble and Garcia-Mila 1992) critical thinking (Kuhn, 1999), argument (Kuhn, 1991; Kuhn, 1992; Kuhn, Shaw and Felton, 1997), problem solving (Kuhn and Phelps, 1992), juror reasoning (Kuhn, Pennington and Leadbeater 1983; Kuhn, Weinstock and Flaton, 1994) and metacognition (Kuhn, 2000; Kuhn and Pearsall, 1998a). She has studied a wide range of age groups from childhood through to adulthood and old age. Like Ennis and Paul, Kuhn sees educational development as a central factor in enabling people to function appropriately in a democratic society. Kuhn is the only author described in this chapter who has carried out empirical research in the area of critical thinking. She uses the abundant research on cognitive development (both her research and others work) to study and teach critical thinking.

Kuhn does not actually offer a definition for critical thinking but she does suggest that any definition should be based on three criteria. Firstly, it should be derived from current empirical research. Secondly, it should be broad enough to allow transfer across a wide range of subject matter. Lastly, it should be embedded in a

developmental framework that can give insight into the origins of cognitive behaviour and their possible directions. She does however use the broad definition put forward by Olson and Astington (1993) in her writings to defend the notion of critical thinking. The definition they propose is, 'Critical thinking is the evaluation of assertions.'

2.7.2 Kuhn's Model of Critical Thinking: Cognitive development frameworks have been used in psychology for many years. Piaget's (1963) 'Theory of Genetic Epistemology' is probably the first, with more recent examples from Perry (1970) and his 'Developmental Scheme', King and Kitchener's (1981) 'Reflective Judgement Model', Kopolowitz's (1987) 'Stages in Adult Cognitive Development' Carroll's (1993) 'Three Stratum Theory of Cognitive Abilities' and Pintrich's (1990, 2000) 'General Framework for self-regulated learning'. Kuhn (1999) also proposes a developmental framework of cognitive abilities in which she includes a developmental model of critical thinking. This is an attempt to merge a cognitive developmental model (predominantly proposed by psychologists) with a critical thinking model (predominantly proposed by philosophers).

Kuhn's (1999) model of critical thinking consists of three forms of meta-knowing or higher order knowing that she suggests makes critical thinking possible, namely; *metacognitive*, *metastrategic* and *epistemological knowing*. Kuhn suggests that the first stage *metacognitive knowing* occurs between the ages of 3-5. The distinguishing marker of movement into this stage is that the child recognizes that people have opinions and beliefs. At this stage they begin to think about thinking as something that humans, including themselves, do. Before this transition Kuhn believes that

children believe there is no inaccurate versions of events, i.e. everything told to them is an exact representation of an external reality. The meta-cognitive form of meta-knowing is essential for critical thinking because it allows the thinker to reflect on the available information because they realise that information is not always fact.

The second type of meta-knowing that Kuhn suggests is required for critical thinking is *metastrategic* knowing. This type of meta-knowing involves the recognition by the person that if someone is engaged in a task like making a decision or constructing reasons they are using strategies of thinking. Kuhn argues that although this process of awareness of other's thinking is absent in under five year olds, it sometimes is poor in older children and adults. Kuhn argues this level of knowledge is crucial in discerning between theory based and evidence based reasoning. Again this skill is found to be poorly developed in some adults, also shown by Anderson and Soden (2001a). Furthermore she states that her research shows that once older participants have a new belief they deny ever having an opposing point of view.

The last development transitions that Kuhn describes are the Levels of *Epistemological Understanding*, which she claims originates in the work of Perry (1970). Within this model there are 4 levels of development namely, *realist*, *absolutist*, *multiplist* and *evaluative*. These levels have an effect on assertions, reality, knowledge and critical thinking. Kuhn argues that these levels of epistemological understanding provide a useful framework by which educators can guide the development of critical thinking skills. Kuhn believes these are the reasons to apply the critical thinking skills when a person develops their metastrategic skills.

2.7.3 Evaluating Kuhn: Kuhn's perspective on critical thinking is different from the other authors in this chapter in that she embeds her critical thinking model in a developmental framework. However parallels can still be drawn between her work and the models of the other major authors. Kuhn's first form of meta-knowing 'metacognition' has been identified by a number of the other authors as an important component of their critical thinking frameworks, for example Ennis and Halpern state that self-regulation and self-checking of one's thinking is an important aspect of critical thinking. Kuhn's second developmental stage of meta-knowing 'metastrategic knowing' is the point where she suggests a person begins to use the strategies of thinking to discern truths. This concept has a number of similarities with critical thinking skills. Kuhn's third stage – 'epistemological knowing' has similarities to the critical thinking dispositions mentioned in the other frameworks. However the links between personal epistemological stances and thinking dispositions may be extremely tenuous and as yet there is no empirical evidence for or against this link.

Overall Kuhn argues that meta-knowing lets a person "know how you know". This puts the thinker in control of their own knowing and as a result puts them in charge of their own lives.

2.8 Merging the Models

This chapter has moved through the individual work of a number of different authors. The first two authors discussed, Ennis and Paul, are firmly based in educational philosophy. Fisher has also worked in this field but also have ties to the disciplines of psychology and mathematics. The last two theorists Halpern and Kuhn

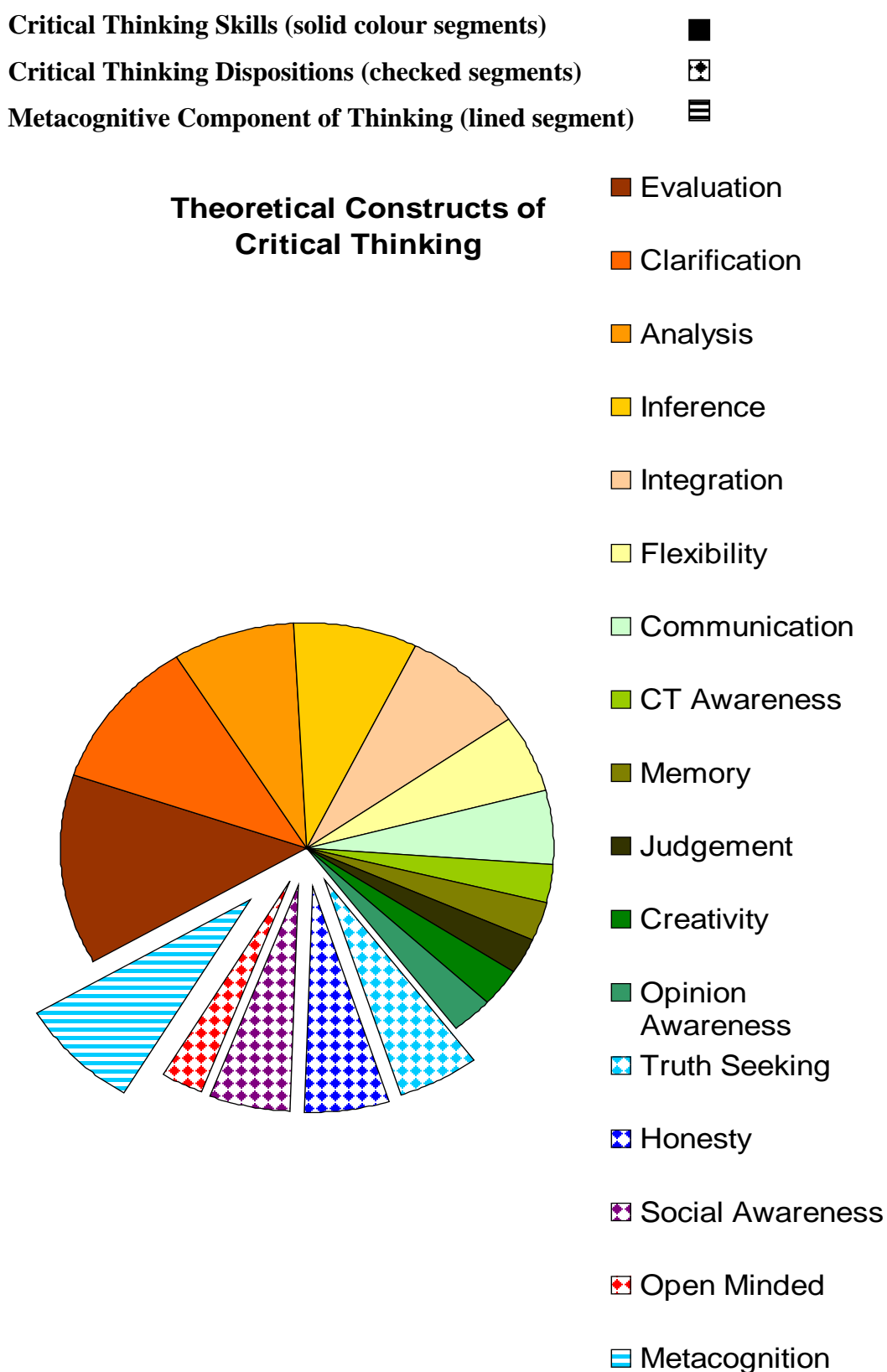
use psychological theory and methods to add to the critical thinking area. Halpern looks at critical thinking from a cognitive psychology viewpoint whereas Kuhn takes the perspective of a developmental psychologist.

The main reason for presenting these author's ideas is to map out the components and concepts within critical thinking. As a result of making explicit the full repertoire of skills, it might be possible to draft a complete set of dispositions and abilities. This draft list is presented in Figure 2.1 at the end of this chapter. There is also a short discussion which accompanies this Figure to aid in its interpretation. Many of the authors suggest that there is an interaction between the components of models; in other words, there is a dynamic or systemic relationship between critical thinking constructs. These relationships are explored in more detail in with empirical findings in Chapter 7 and in the general discussion Chapter 8.

Figure 2.1 below can be broken into three sections; the first section represents critical thinking skills (solid coloured sections of the chart), the second displays critical thinking dispositions (check pattern sections) and the third shows meta-cognitive components of critical thinking (line patterned sections). The constructs and areas represented by those constructs are derived from the matrix in Appendix 2.1. This matrix has compared and contrasted the critical thinking models of the major authors in the field as discussed in this chapter. This matrix is subjective in that it draws their opinions together under general headings. These opinions and ideas do not fit perfectly but if the components seemed to have similar characteristics they were drawn together. The reasoning behind the chart is to get a general theoretical model of critical thinking based upon the opinions of these major authors.

Evaluation, Clarification, Analysis and Inference are critical thinking skills mentioned by most of the authors and therefore they account for a large proportion of the critical thinking 'picture'. Integration and flexibility are related skills in that they both describe the dynamic use of critical thinking skills. Again these two skills were identified by most of the authors. The authors are less unanimous in their opinion on the remaining skills. These skills tend to be ones identified by the cognitive psychologists, Halpern and Kuhn, e.g. decision making and memory. However, some of the philosophers do concur with some of these skills. For example, Ennis sees communication as an important sub-skill of clarification and Fisher agrees with Halpern on the inclusion of Creativity. Critical thinking dispositions are not as prominent in the authors' models as critical thinking skills. Ennis and Paul are explicitly the main proponents of critical thinking dispositions however dispositions are also identified by Halpern and Kuhn's developmental category, Epistemological Knowing, has many related aspects. It is suggested that the four groups in the chart below, Truth-seeking, Honesty, Social Awareness and Open-mindedness offer the best summary of the components offered by the theorists. Lastly Metacognitive components are suggested by a number of theorists namely; Ennis, Halpern and Kuhn. The authors identified components of this construct (see Appendix 2.1) seem to be influenced by both critical thinking skills and dispositions.

Figure 2.1: Amalgamation of the critical thinking models proposed by the major authors and identification of theoretical constructs of critical thinking.



2.9 Comparing the Merged Model with the Delphi Analysis

The main difference between the conclusions in Figure 2.1 and the Delphi Analysis is that a meta-cognitive dimension which is not identified in the Delphi report features in a number of the main theorists' frameworks. Nevertheless, the Delphi Report identifies a critical thinking skill of self-regulation which has many of the connotations of meta-cognition like self-examination and self-correction. Furthermore, although the majority of skills and dispositions are repeated in both the Delphi Report and Figure 2.1, there are some exceptions. For example, skills of communication and clarification do not get substantial treatment in the Delphi Report but are mentioned frequently by the five theorists. Furthermore, skills associated with the study of cognitive psychology like memory, creativity and judgment are not included in the Delphi Report. Likewise, the skill of being able to integrate critical thinking skills to suit particular objectives is mentioned by a number of the critical thinking writers but not explicitly outlined in the Delphi Report. Conversely, critical thinking dispositions get a more thorough treatment in the Delphi Report. The report identifies nineteen dispositions, while dispositions identified in the major frameworks can be summarised under four headings. In conclusion, the Delphi Report seems to emphasise critical thinking dispositions more heavily while the frameworks tend to concentrate on critical thinking skills. These differences are probably a result of the composition of the two groups used to produce these two opinion summaries. That is, the majority of Delphi participants were philosophers. Regardless of the differences between these two forms of summary, there is a high degree of overlap and it is these overlapping constructs that are of particular pertinence when attempting to assess critical thinking.

Chapter 3

The Assessment and Measurement of Critical Thinking in Higher Education

3.1 Introduction

The main theme of this thesis concerns the *measurement* of critical thinking. The following chapter discusses the measurement of critical thinking in the specific context of higher education. Initially the chapter highlights why critical thinking measurement is important in higher education, it also discusses the two main types of measurement, namely, domain-independent and domain-specific assessment and it reviews some measurement tests and assessment techniques that are used to measure the critical thinking of tertiary level students. Appendices 3.1 and 3.2 summarise the reliability and validity information that is used in this chapter.

3.2. Critical Thinking Assessment Issues

3.2.1 Introduction

There are a number of methods currently being used to measure critical thinking but before one chooses a method, the assessor must be clear why s/he wants to measure critical thinking. The following Section 3.2.1 details the reasons why one might want to assess critical thinking. Once the reason for assessment is established the method of assessment must be chosen and the two main methods of critical thinking assessment are presented in Section 3.2.2, measuring critical thinking as embedded in knowledge domains and measuring critical thinking as a general reasoning skill. Finally, when these two decisions are made the assessor can select an appropriate methodology. Examples of psychometric approaches to measuring critical thinking are described in Section 3.3.

3.2.2 Why is Critical Thinking Assessment Important in Higher Education?

The following section explores four major reasons for measuring critical thinking in higher education settings, namely ‘degree accreditation’, ‘assessing standards’ ‘evaluating critical thinking interventions’ and ‘selection’. These purposes for critical thinking assessment differ in a number of ways, i.e. their frequency of use, the assessment method that is suitable and even their level of controversy. The following paragraphs elucidate some of the issues surrounding these purposes.

Firstly, critical thinking is an important criterion to be used when ‘accrediting’ educational programmes or courses. Many educators have suggested the development of critical thinking is particularly important in the higher education process (Barnes, 1992; Blaine-Carpenter and Doig, 1988; Cromwell 1992, 1993; Giancarlo and Facione, 2001; Walker and Finney, 1999). Recent exercises such as higher education benchmarking (HMSO, 1997) in the UK have highlighted the need for university courses to promote key and transferable skills such as critical thinking. *The Dearing Report* (HMSO, 1997) suggested that critical thinking development is a distinctive feature of higher education compared to other levels of education. Furthermore, it is suggested that exposure to the university experience promotes critical thinking and this assertion is supported by research looking at experience of university life and its effect on critical thinking development; factors explored include campus culture (Tsui, 2000), faculty staff’s attitudes (Tsui, 2001) and diversity of experience (Pascarella, Palmer, Moye and Pierson, 2001). In short, a large body of opinion and evidence identifies the crucial role higher education has in promoting the development of critical thinking. Therefore, accreditation criteria

increasingly pay attention to whether critical thinking is included in the learning outcomes for graduates and the teaching methods that are adopted to promote critical thinking in students.

Secondly, critical thinking is an important criterion for assessing standards and for awarding a specific degree level, e.g. 1st class honour, 2.1 or 2.2. The following quotation is from the British Psychological Society's (2003) guidance for marking students' written work: 'The very best candidates will demonstrate considerable breadth of reading, an ability to integrate material from different parts of the syllabus in a convincing way, a capacity to analyse and *critically evaluate* both theory and the results of empirical investigations and an ability to make a personal contribution to the development of the topic under discussion.' This statement from the BPS emphasises critical thinking, evaluation and the ability to integrate as skills to be demonstrated when setting standards for awarding the best degrees to students.

Because of the importance of critical thinking in higher education, many institutions run courses specifically designed to improve students' critical thinking. Despite criticisms by McPeck (1990) on the value of 'stand alone' critical thinking modules and courses, this type of programme is popular in a variety of disciplines in a number of universities throughout the world, e.g. California State University - Psychology; Dublin Business School - Philosophy; National University of Singapore - Architecture; New York Institute of Technology; Trinity College Dublin – In their 'Broad Curriculum', i.e. courses across disciplines; University College Dublin – Psychology, University of Alberta - Psychology; University of Liverpool - Medicine and many more. These interventions have varied durations, from a single class

through to a number of modules. Furthermore, explicit instruction in critical thinking can even be the central ethos of the whole university curriculum, which is the case in a number of Liberal arts colleges in the US, e.g. Alverno College – Wisconsin USA. Interventions also range in content and style from developing metacognitive skills and problem solving (Hanley, 1995) to reasoning and peer interaction (Anderson and Soden, 2001). Therefore, the need arises to assess the effectiveness of these critical thinking pedagogies using pre/post research designs and assessment tools that focus on forms of critical thinking (Anderson and Soden, 2001; Conklin, 1987; French and French, 1991; Hanley 1995; Reed and Kromery, 2001; Tsui 1999 and 2002, Van Gelder, 2001). In these contexts there is an increasing demand for reliable and valid psychometric tests of critical thinking.

The final purpose for critical thinking measurement described here is probably the least used and most controversial, i.e., for academic selection. Job selection using psychometric tests is a method used by occupational psychologists since the Second World War. There is growing discussion on the use of psychometric testing for selection to degree programmes both in the US (Kreiter, Stansfield, James and Solow, 2003), UK (Searle and McHarg, 2003) and China (Higgins and Sun, 2002), and this is becoming particularly popular for high demand courses like medicine (Higgins and Sun, 2002; Kreiter et al, 2003; Searle and McHarg, 2003).

Caution is often advised with high stakes psychometric testing for jobs or education courses (Cooper, 1998) as these tests may not be sufficiently reliable and valid. However, the selection techniques currently used in higher education in the UK (mainly A-levels) also have a number of problems. These difficulties include;

cultural bias against people of a non-UK background, socio-economic bias, subjective marking, administration errors, unequal skill development across subjects (McEwen, McGuinness and Knipe, 2001) and emphasis on knowledge rather than thinking (Richardson, 1994; McEwen, McGuinness and Knipe, 2001). The US system uses standardised alphanumeric reasoning tests, which specifically select for college entry (SAT and GRE) and these are combined with a test of academic attainment (grade point average GPA) for selection purposes. Both these types of selection tests have been shown to be predictive of college attainment, with the GPA showing greater strength of prediction (Astin, 1997). Lastly, as higher education invites a lifestyle change as well as a pedagogical change, personality variables may have some predictive validity of a student's success at university or suitability for a certain discipline within university. Personality is not considered in current selection models for higher education but again it has been suggested (Searle and McHarg, 2003). Many of these selection issues are on the agenda of a steering group (Admissions to Higher Education Steering Group (AHESG), 2003) investigating higher education selection in the UK.

3.2.3 Assessing Critical Thinking in Knowledge Domains vs. General Skills

The assessment of critical thinking raises more general questions about the nature of knowledge and reasoning. On the one hand, thinking and reasoning can be seen as a general cognitive processing ability that is readily transferable across different topics and contexts. On the other hand, thinking and reasoning can be seen as highly embedded in knowledge and disciplinary contexts such that it is only worthwhile assessing critical thinking as it relates to particular knowledge areas (e.g. History, Mathematics and Art). Traditional university assessment tends to be associated with

the embedded position while the psychometric measurement of critical thinking reflects the general cognitive processing view.

Assessment of critical thinking at university normally occurs when an academic analyses a student's written work on a subject specific topic for evidence of critical thinking. Difficulties encountered with this method of assessment are that students often see written work as an opportunity to show how much information s/he has acquired about a particular subject rather than as an opportunity to demonstrate critical thought. Additionally, it is often difficult for students to judge how much weight will be given to critical thinking versus knowledge in a marking scheme. Research that specifically analyses students' argument skills in essays shows that they rarely use their knowledge in an evaluative or critical manner (Anderson, Howe, Soden, Halliday and Low, 2001).

The main difference when assessing critical thinking as a general skill is that subject knowledge is no longer the focus of attention - only the critical thinking. Critical thinking tests of this kind rely on 'general knowledge' or controversial issues as prompts for critical analysis, critical argument and critical evaluation. Tests that measure critical thinking as a general skill can range from essay tests where the student has to construct a critical argument on a newly introduced topic, through to multiple choice tests, where students are asked to choose from a range of options that show their ability to recognise assumptions, draw inferences and evaluate arguments. There are also short free response type tests for measuring general critical thinking skills, where, for example students have to state assumptions made in a passage or recognise conclusions reached in that passage. The grading of critical thinking essay

and short free response tests is carried out in a similar way to the assessment of critical thinking within knowledge domains. In both cases the rater has to assess the student's writing for evidence of critical thinking. However, the raters should only be looking for evidence of critical thinking and not evidence of subject knowledge as they would normally do when assessing students' university work. Although multiple choice tests are more easily scored, Ennis (1996a) suggests further disadvantages specific to multiple choice tests, in that they may not be comprehensive enough to measure the full scope of critical thinking and differences in beliefs of the test producer and students taking the test can be detrimental to the student's score.

The development of critical thinking tests has a long history but they are relatively few in number and have predominantly been produced in the US. Also, many of the test writers are associated with the educational philosophy tradition in critical thinking and the tests have not been subjected to traditional psychometric scrutiny.

3.3 Critical Thinking Measurement Tools

3.3.1 Introduction

The following sections describe a number of critical thinking tests. Tests reviewed include the Watson-Glaser Critical Thinking Appraisal, The Cornell Critical Thinking Test, The Ennis-Weir Critical Thinking Essay Test, The California Critical Thinking Skills Test and The California Critical Thinking Disposition Inventory. A brief description of the structure, content and administration is provided for each of these tests. Furthermore, some details about the background and academic work of each of authors of these tests are included. The reason for reviewing these tests is to

provide a context for comparison with the two tests (CCTST and CCTDI) used in the thesis.

3.3.2 Watson Glaser Critical Thinking Appraisal, WGCTA

Watson and Glaser (1964) were the first to attempt to explicitly measure the concept of critical thinking. Watson's work in the area stretches back to 1925. The WGCTA is a much used measure of Critical Thinking (Loo and Thorpe, 1999). It has been used prolifically by occupational psychologists for the selection and promotion of candidates into management positions (Watson and Glaser, 1991). There are a number of forms of the Watson-Glaser Critical Thinking Appraisal (WGCTA), form Ym and Zm (1964), A and B (1980), C (1991) and S (1994). The original forms Ym and Zm were refined in 1980 by removing items that contained gender stereotypes or dated information to create Forms A and B. Form C is a revised version of the Form B test adjusted to suit UK populations. In this Anglicisation, 19 items were unchanged, 20 items received slight alteration, 36 had minor amendments and 5 underwent major changes (Watson and Glaser 1991). Form S is a recently revised version of Form B which is shorter, featuring 40 items compared to Forms A, B and C which have 80 items (see Geisinger, 1998 for a psychometric review of Form S).

Form C is described here because it is the version most suitable for use in a UK population. It is suitable for use with adults with a post 16 year education. There are 5 subscales namely; Inference, Recognition of Assumptions, Deduction, Interpretation and Evaluation of Arguments. The test can be group administered and takes approximately 30-40 minutes and can be carried out with or without a time limit.

The WGCTA test manual contains a careful, well constructed definition of critical thinking and the test items are generally derived from this definition (Helmstadter, 1985). The WGCTA has had prolific use over the last 40 years and thus there is abundant detail on its reliability and validity (see Appendices 3.1 and 3.2). The majority of items in the WGCTA focus on political or social issues like war, disease, smoking and astrology and there is little use of knowledge information like science or literature. The authors also divide the item issues into neutral and controversial with the controversial being assumed to provoke a more rigid stance. However, the division of items into this distinction seems unproductive as there appears to be a sliding scale of issues rather than two distinct groups (Berger, 1985). A further problem with the items is that, in most cases there are only two possible responses (except for the inference sub-scale) and despite this there is no means of correction for guessing.

Overall, there are a number of problems with the validity of the WGCTA. The first main problem with this test is pointed out by Helmstadter (1985) in his review of the test in the *Mental Measurements Yearbook*. He suggests that sometimes personality is the main influence on an item response rather than the person's ability as a critical thinker, especially in the inference sub-scale. This problem could affect the standardised nature of scoring of the WGCTA. The second major problem is that the manual provides evidence of concurrent validity with mental ability tests and tests of scholastic aptitude but there is no evidence of correlations with similar critical thinking tests (Helmstadter, 1985) such as the Cornell Critical Thinking Tests (Ennis and Millman, 1985). However, the manual of the California Critical Thinking Skills

Test (Facione, 1991) shows a significant correlation between the two tests (see Appendix 3.1). Generally, the WGCTA shows significant concurrent correlations with alphanumeric ability tests like the SATs or ACTs but lower correlations with tests of scholastic achievement like GPAs or university grades (see Appendix 3.1).

Overall the levels of WGCTA reliability seem to be acceptable (see Appendix 3.2) and these estimates remain high regardless of method used (KR-20, Alpha or split half, see Appendix 3.2). However, this is not demonstrated in a test manual which only gives estimates of split half reliability. Watson and Glaser (1991) do not recommend the use of sub-scales within the test because of the relatively low numbers of items in each sub-component, except for their use for identifying weak areas in a class or group. This hints to reliability and validity problems of these sub-scales but this is not explicitly stated in the manual and no evidence is presented on subscale reliability or validity.

The test manual for the WGCTA version C presents a comprehensive, well differentiated set of norm tables. These norm tables are available for groups including sixth form students, university/polytechnic students (MBA and Business) and promotion candidates (Price Waterhouse Cooper, Police force). The manual also features 33 norm tables with version A and B though most feature US populations (Watson and Glaser, 1991).

The authors urge caution about using the test as a stand-alone measure for high stakes testing such as individual selection for jobs, promotion or education but they do suggest that it can suffice as an initial screening tool. Overall, the WGCTA is a

global measure of critical thinking even though it has named sub-scales. The test validity may be confounded by assessing some element of critical thinking dispositions without separating them from abilities.

3.3.3 Cornell Critical Thinking Test Level Z, CCTT (Ennis and Millman 1985)

Ennis has developed two tests of critical thinking. One is a multiple choice test (Cornell Critical Thinking Test) and the other is an essay type test (Ennis Weir Critical Thinking Essay Test). The Cornell Critical Thinking Test Level Z was designed for exceptional high school students, college students and adults compared to the Cornell Critical Thinking Test Level X which is designed for lower ability groups, i.e., American grades 4-14. However the distinction between who should take level Z and level X is not well defined in the manual (Malcolm, 1992). This test is broken down in separate sub sections which assess 'induction, credibility, prediction and experimental planning, fallacies (especially equivocation), deduction, definition and assumption identification', (Ennis, 1992, p. 81). There is 50 minute time limit for this test, which should be administered in a relaxed unthreatening environment to optimise participant performance (Malcolm, 1992). The 1985 version of this test is similar to the 1971 version with a few minor amendments to both the test and manual (Hughes, 1992). The test has 52 items which are broken down into sub-scales but there is no factor analytic evidence that these sub scales have construct validity (Hughes, 1992; Malcolm 1992). This point would suggest that the test was constructed and scaled using content or face validity based on Ennis's framework of critical thinking (see Section 2.3.2). Looking at Appendix 3.1 it can be seen that the CCTT correlates with scholastic ability but the range is so wide that no firm conclusions can be drawn.

The reliability of the test is adequate (see Appendix 3.2) for low stakes testing but does not in any incidence rise above .80. Also there is no evidence given for the sub-scale reliabilities; so the identification of sub-scales may be unwarranted. It is assumed their reliability is very low due to the overall reliability being approximately .70 and the fact that there are seven sub-scales. Another negative point about the test manual is the poorly described norms. These norms are too general in their nature and do not give figures for specific groups like age or level of education (Hughes, 1992; Malcolm, 1992). One positive point about the test is that it is simple to use, and to score, and is well explained.

The authors suggest the best use of the test is in roles similar to some of those discussed in Section 3.3.2, i.e. for evaluating critical thinking interventions and degree programmes for critical thinking development. They even suggest that the tool could be used as a teaching aid as there are lesson prompts in the form of item answer rationales (Malcolm, 1992). However, the authors and others suggest this test should not be used for decisions concerning an individual, i.e. selection, especially when used on its own (Hughes, 1992; Malcolm, 1992). In conclusion, the CCTT is best used as a low stakes educational tool.

3.3.4 Ennis Weir Critical Thinking Essay Test, *EWCTET* (Ennis and Weir 1985)

This test is aimed at the range of groups from adolescents to college students. There are no explicit subscales within this test as it is in an essay response format but there

are a number of competence areas identified in the manual namely ‘getting the point, seeing reasons and assumptions, stating one’s point, offering good reasons, seeing other possibilities, responding appropriately to and/or avoiding equivocation, irrelevance, circularity, reversal of an if-then relationship, straw-person fallacy, overgeneralization, excessive scepticism, credibility problems, and the use of emotive language to persuade’, (Ennis and Weir, 1985, p.2). The test takes 40 minutes to administer with 10 minutes recommended for reading source material and 30 minutes for writing an essay response. The authors recommend two uses of this test, namely, the assessment of critical thinking ability and as a teaching tool. The authors place emphasis on its educational uses and suggest that it can be used by both Socratic and Didactic teachers (Poteet, 1989). Socratic teachers use student centred discussion as their teaching method and can employ the participant test responses as a dialogue stimulus for a class within a critical thinking module (Tompkins, 1989). Didactic teachers who employ teacher centred instruction as their pedagogical method could use the student responses in diagnostic way; so they can identify shortfalls in critical thinking development in their classes and respond with the appropriate tuition. The teaching emphasis of this test is furthered by Ennis and Weir’s written permission to reproduce the tests and test materials for use within a class. The main task in the test is to write a critically thought out letter to the editor of a newspaper in response to 8 logically flawed letters on a real world issue. The topic for discussion in the EWCTET is traffic problems within a fictitious city. The exercise ‘is intended to help evaluate a person’s ability to appraise an argument and to formulate in writing an argument in response, thus recognising a creative dimension in critical thinking ability’ (Ennis and Weir, 1985, p.3). There is a well structured and detailed scoring key for the EWCTET with a systematic method of

constructing a total score. However, there are no norms provided by which to compare total scores. The authors recommend that the test be scored by a person who has at least college level experience of informal logic or critical thinking. Furthermore, the grader is advised to consider thinking ability rather than writing ability when marking student responses.

The manual presents no evidence of construct, predictive or concurrent validity. This is unfortunate because some concurrent validity of the test could easily be obtained by correlating results with other tests of critical thinking (Tompkins, 1989). However, there is a suggestion that the test has content validity because the task is scored on critical thinking ability according to Ennis and Weir's (1985) conceptualisation (Poteet, 1989). There are two inter-rater reliability estimates given in the manual (see Appendix 3.2) but these are based small samples numbering 27 and 28.

In conclusion, the authors have produced a well described, open ended test of critical thinking, which could be very useful as a teaching aid. However, the test requires more reliability and validity checks before it is used for assessing standards, accreditation or formal selection.

3.3.5 The California Critical Thinking Skills Test, *CCTST*, Form A (Facione, Facione, Blohm, Howard and Giancarlo, 1998).

This test can be administered to third level students and adults. Facione et al (1998) have produced a multiple choice test that targets core critical thinking skills as

outlined in the 'Delphi Report'. The major sub-components of critical thinking suggested by the Delphi Report were Analysis, Evaluation, Inference, Interpretation, Self-regulation and Explanation. However the test purports to measure Analysis, Evaluation and Inference with three correspondingly named sub-scales. Participants have 45 minutes to complete the 34 items in the test. Facione et al outline the main uses of the test are as a critical thinking program evaluation tool, identification of group or individual critical thinking weaknesses or individual selection. These uses are similar to those of the WGCTA (Section 3.3.2) and similarly the author warns against using the test as the sole measure for occupational or educational selection.

Facione et al. (1998) state that the test was produced from a bank of 200 items that they have developed during 20 years of research into critical thinking. He also states that the incorrect answers to the multiple choice items represent commonly seen mistakes when using critical thinking or reasoning, e.g. choosing a particular incorrect response because of a personal bias. However, test reviewers (McMorris 1995; Michael, 1995) have stated concerns on the selection and placement of some of the items in this test. They suggest that in a few cases the best answer is not the one outlined in the marking scheme as the correct answer. The reviewers' (McMorris, 1995) concern is compounded by the lack of information around expert involvement in item placement in the sub-scales. Two important questions can be raised about the structure of this test. What were the criteria used for placing the item within a particular sub-scale? What were the criteria employed for item retention? These questions are looked at in more detail in Chapter 6.

This test has good content and face validity due to its construction around the 'Delphi Report' (McMorris, 1995) and the items purport to measure theoretically derived critical thinking skills like inference, analysis and evaluation. The concurrent validity information does not cover as many groups as the Watson-Glaser but the manual does provide a number of validation checks with a range of student groups (see Appendix 3.1). Critical thinking research and assessment have become popular and as a result so has the CCTST, especially as there are so few critical thinking tests available. The manual provides correlation information with the CCTDI, Watson-Glaser, US university entrance exams (SAT, GRE and ACT) and age. The CCTST has strong relationships with the alphanumeric reasoning tests used in US entrance exams (especially the GRE) and the WGCTA with a lower but significant relationship with the CCTDI. However, there seems to be no relationship between age and score on the CCTST (see appendix 3.1). Correlations with the other measures of critical thinking and intelligence would be helpful as it would be useful to see if relationships with intelligence tests were weaker than those with other critical thinking measures as this would show critical thinking uniqueness as a concept. There are a number of problems concerning the CCTST's construct validity (McMorris 1995; Michael, 1995). Facione et al. (1998) provide substantial psychometric information in the manual related to construct validity but neglect to mention factor analysis as a way of discerning construct validity. One way Facione et al. claim construct validity of the test can be assessed is to measure students before and after a critical thinking course but no evidence is reported. Overall, the general view is that more psychometric analysis of this tool is required to provide construct validity information (Michael, 1995).

Reliability estimates of the test are not as high as is normally required of a test that may be used in high stakes testing (see Appendix 3.2). However, Facione argues that the lower Kuder Richardson – 20 reliability estimates are acceptable for two reasons. Firstly, the items are dichotomous, i.e. they are either correct or incorrect and as a result inter-item correlations will be lower (Nunally, 1978). Secondly, Facione believes that the test is not measuring a single homogenous ability but rather the individual components which make up critical thinking namely, evaluation, inference and analysis. Therefore there is no reason why items measuring different skills should have high inter-item correlations. If Facione is correct in his assertion of separate skills then at least his component scales should have a high internal consistency yet this information is not provided in the test manual. In addition, his argument raises doubts about the meaning of total score on the CCTST.

There are 3 groups of norms provided in the manual, junior college students (N=781), nursing students (N=153) and police cadets (N=224). These norms are not of a wide range of groups and therefore generalisability could be problematic. The lack of norms is surprising considering the fact that a Spanish and Chinese version of the test is available. The authors suggest that local population norms may need to be obtained, especially for specialist groups or cross cultural studies.

Overall the test has a good theoretical basis due to its partial overlap with the Delphi report. However there are problems with its validity, provided norms and the specificity of its reliability estimates. More research into the test and closer scrutiny of the items may yield a more promising test of critical thinking. These tasks were carried out as part of this thesis and are described in Chapter 6. Since the work on

this thesis began Facione has published a new version of the CCTST Form A called the “California Critical Thinking Skills Test 2000” (Facione, 2000).

3.3.6 The California Critical Thinking Disposition Inventory (CCTDI) (Facione, Facione and Giancarlo, 2000).

This test is suitable for use with the following groups, secondary students, college students and adults. There are seven sub-scales in the test and a total of seventy-five items. The sub-scale names are; Truth Seeking (12 items), Self-Confidence (9), Systematicity (11), Analyticity (11), Maturity (10), Open-Mindedness (12), Inquisitiveness (10). Like the CCTST the authors claim this inventory is suitable for group or individual testing. There are 75 items in the test and takes approximately 15 minutes to complete.

The CCTDI is also based on the majority consensus reached during the Delphi exercise. The purpose of the test is to measure the participant’s disposition or attitude towards using critical thinking. As the authors state in the manual, the test does not measure critical thinking abilities but an affective component of critical thinking and they provide good descriptions of these affective components at the beginning of the manual (Ochoa, 1995). These sub-scales are elaborated on in Chapter 5.

The uses of the CCTDI are similar to that of the CCTST in that it can be used to evaluate courses or critical thinking programmes for their ability to develop critical thinking dispositions. Furthermore, the test can be used diagnostically to identify problem areas in the dispositions in specific groups and possibly on an individual basis. A related use is research in the identification of the Kuhnian like stages of

development epistemological beliefs (see Chapter 6). Facione and others (Callahan, 1995; Ochoa, 1995) urge caution for using the CCTDI for high stakes selection.

Again, like the CCTST, the CCTDI has a good content validity due to its origins embedded in the expert opinion resulting from the Delphi exercise. Furthermore information is provided on how the Delphi Report was used to produce the items that appear on the scale. The authors wrote 250 items based on each of the phrases that the Delphi consensus had decided were the characteristics of 'an ideal critical thinker'. There is also information provided on the selection criteria for the eventual test items. The authors reduced the 250 statements to 150 by having them reviewed for repetition and ambiguity by college level critical thinking educators. The final 75 items were chosen on the statistical strength of their internal consistency and discriminating power although no additional numerical evidence for this is provided in the manual. A factor analysis is partially reported to provide evidence of construct validity of the sub-scale. Only the mean factor loading/item is given for each subscale and not loadings for each item. This is insufficient information, as some items may have had very high and acceptable loadings while others may have had very low loadings and should have been removed from the test. The concurrent validity information of the test provided in the manual is based on correlations with two widely investigated and, empirically sound constructs namely Openness to Experience from the NEO PI-R (Costa and McCrae, 1991) and Ego Resiliency (Giancarlo, 1993). There are no correlations provided with other explicit measures of critical thinking dispositions because no similar critical thinking dispositions measures are currently available. However, recently Facione and colleagues have produced 'The California Measure of Mental Motivation' (Facione and Giancarlo,

1999) which purports to measure critical thinking constructs namely creative problem solving, learning orientation, mental focus and cognitive integrity. Further research on both scales may provide additional concurrent validity.

Generally, the overall reliability of the CCTDI is quite impressive (see Appendix 3.2) but the test has 75 items and a large number of items has the effect of increasing Cronbach's Alpha. The sub-scale estimates are still quite high but these estimates were obtained from a poorly described sample restricted to American college students and may have some replication and generalizability difficulties (see Appendix 3.2 and Chapter 5).

The authors provide no norms for comparison in the manual; instead they suggest arbitrary figures for interpreting a participant's score. Each sub-scale has a maximum score of 60 and, overall, the maximum score is 420. The manual states that the cut score for each sub-scale is 40, with those scoring over 50 showing strength in that particular disposition, and anyone showing a score below 30 displaying a weakness in that disposition, with those between 30 and 40 being ambivalent with regard to the disposition. Overall, the ambivalent range is 210-280, above and below this range shows overall strength and weakness in critical thinking dispositions.

This test is available in seven languages, English, Finnish, French, Hebrew, Japanese, Portuguese and Spanish. However the lack of norms and limited psychometric refinement could make its cross-cultural transferability suspect. Nevertheless, it is based on a solid theoretical base and could be quite useful if a

researcher is willing to recalibrate it in order to improve its stability (Callahan, 1995). A recalibration of this type is carried out and described in Chapter 5.

3.4 Conclusion

There are number of points of note to be taken from the chapter. The first is that a psychometric instrument should have solid foundations in psychological theory. The tests reviewed in the chapter have been based on many years of research (WGCTA, CCTT, EWCTET) or large-scale expert consensus (CCTST and CCTDI) and have good content validity. Secondly, test scores should show good reliability and validity information for use particularly when these uses are high stakes. Thirdly, when developing a psychometric test it is crucial to be explicit on which test scores can be used and that these scores have sufficient reliability and validity. For example, should one use overall scores or do sub-scales scores have adequate reliability and validity for individual use. The tests above do not have sufficient psychometric information for a user to make these decisions with conviction.

From the review of the five tests it is apparent that none are completely satisfactory from a psychometric point of view. However the California tests do have some important advantages over the others. They are the most recently constructed. They grew out of the Delphi consensus and cover both critical thinking skills and dispositions, which were identified as important constructs in the merged model outlined in Chapter 2. Despite their psychometric naivety it was decided to select them as the instruments for the empirical work in the thesis. The remaining chapters show how the Californian tests performed when used with student samples in a UK

higher education context. A substantial proportion of the work focused on strengthening the psychometric properties of the tests.

Chapter 4

Rationale and General Design

4.1 Introduction

The purpose of this thesis is to examine critical thinking in undergraduate students. General questions to be addressed include: What are the components of critical thinking? Can these components be reliably and validly measured? Can these component measurements be used to discriminate between UK student groups and predict educational outcomes? In essence, the thesis defines what is being measured, describes how best to measure it and shows what can be done with these measurements.

The information for addressing the core questions stated above is contained in both the literature review and empirical studies of the thesis. The following chapter acts as a bridge between these two major sections and explains the rationale and general structure of the thesis. The chapter begins by re-stating the main points from the previous research reviewed in the last three chapters (Section 4.2). It then details how the conclusions link to the empirical part of the thesis and what specific research questions have emerged as important (Section 4.3). Finally, it details the general design and participant information for the studies which follow in the next three chapters.

4.2 Review of Introductory Chapters

Chapter 1 provided a broad overview of critical thinking and described the evolution of critical thinking theory into its present state and current applications. It frames the thesis topic by detailing perspectives of the people interested in critical thinking and by describing the history of the area. The chapter allows the reader to position the thesis in the wider field of critical thinking research.

Chapter 2 concentrated on the major theoretical frameworks put forward by authors who have conceptualised both critical thinking skills and dispositions. At the end of the chapter, the core concepts and arguments of the different theorists were drawn together, the main overlaps and similarities in the meanings of their constructs were identified and mapped onto a diagram (Figure 2.1) that shows, at a glance, the kinds of concepts that re-occur across the different theoretical positions. This overview also pointed to what constructs might be worthwhile measuring, at least with regard to the perceived importance in the different theories. The conclusion also provides some detail about the differences between the framework summary at the end of this chapter and the other major opinion survey available, i.e., the Delphi Report.

Having identified what might be worthwhile measuring, the goal of Chapter 3 was to explore the best available methods of measuring it. This process began by detailing the major tools used for measuring critical thinking skills and dispositions. The two tools that were considered best for mapping onto the derived critical thinking framework were the California Critical Thinking Dispositions Inventory (CCTDI) and the California Critical Thinking Skills Test (CCTST) because they claim to

measure two of the major constructs outlined in Figure 2.1, i.e., critical thinking skills and dispositions.

4.3 Research Questions and Structure of the Empirical Studies

Although the CCTDI and CCTST were found to best map onto the general map of critical thinking constructs (Figure 2.1), there were still problems apparent with their reliability and validity. Therefore, the initial empirical work in the thesis is psychometric. Chapters 5 and 6 research the issues of the strength of the psychometric properties of the instruments. The remaining empirical chapter, Chapter 7, investigates the predictive strength of the two tests and also outlines the best model of measures for predicting degree outcome.

Research Question 1

What is the construct validity of the CCTDI for a UK undergraduate population?
(Chapter 5)

Both the CCTDI and CCTST were developed in the US and their constructs and factor structures were validated on US populations. Their subsequent use had mainly been with US samples, although not exclusively. The previous review (section 3.3.6) suggested that the psychometrics of the CCTDI were questionable. As this research was the first time the inventory was used on a UK sample, the first step was to investigate the reliability and validity of the CCTDI. Initially, it was intended that this would form only a small part of the overall empirical work of the thesis. However, as will become clear in Chapters 5 and 6, this turned out not to be the case and additional research questions were generated. A series of confirmatory and

exploratory factor analyses (over three studies) resulted in an inventory with substantially the same number of subscales, but with a reduced number of items overall and substantial item movement between the sub-scales, called the CCTDI – UK - to distinguish it from the original inventory. Comparisons were made between the factor structures of the CCTDI and CCTDI-UK. The UK version was then subsequently used in the remaining studies.

Research Question 2

Is there evidence of convergent and group contrast validity for the CCTDI-UK?
(Chapter 5)

Having settled on the new 6 sub-scale structure of critical thinking dispositions, additional questions about construct validity were then addressed in three further studies. Convergent validity was checked by comparing factor subscale scores with each other and a sub-scale from a more established personality inventory that measures related constructs – the openness to experience subscale of the NEO PI-R (Costa and McCrea, 1991).

The group contrast validity question was examined in a longitudinal study by comparing differences between scores of undergraduate students who completed the inventory at three time-points (1st year, 2nd year and 3rd year) as they progressed through their undergraduate programme and by a cross-sectional study that compared critical dispositional scores from a 1st year ‘entry’ sample with a 3rd year ‘exit’ sample.

During the completion of the convergent validity checks on the subscales, it emerged that there was a higher-order structure to the critical thinking disposition factors. However, these higher order factors proved subsequently to be less robust than individual sub-scales when testing predictive models (see Chapter 7).

Research Question 3

What is the reliability of the CCTST in a UK undergraduate student population?
(Chapter 6)

The purpose of this question was essentially similar to the one addressed for the critical dispositions inventory. This is the first time that the CCTST has been used on a UK population. Sub-scales on the CCTST were checked for internal reliability, items were removed, and reliabilities were re-checked. A modified test, called the CCTST-UK, was produced and the psychometric properties were compared with the original CCTST. CCTST-UK was then used in all subsequent studies in the thesis.

Research Question 4

Is there evidence of convergent and group contrast validity for the CCTST-UK?
(Chapter 6)

Convergent and group contrast validity of the CCTST-UK were examined in two studies. Convergent validity was checked by examining total scores and sub-scale scores with measures of non-verbal intelligence (Ravens Advanced Progressive Matrices Short Form - RAPMsf), and academic achievement (A-levels, 3rd year

marks, and overall degree average which is a combination of 2nd year and 3rd year marks).

The group contrast validity was examined in a cross-sectional study that compared differences between total CCTST-UK scores and sub-scale scores from a 1st year 'entry' sample with a 3rd year 'exit sample of undergraduate students.

Research Question 5

Do critical thinking dispositions predict additional variance in degree outcome beyond that predicted by A-level marks? (Chapter 7)

Research Question 6

Do critical thinking skills predict additional variance in degree outcome beyond that predicted by A-level marks? (Chapter 7)

The previous research questions addressed issues about the psychometric properties of the revised measures for assessing critical thinking dispositions and critical thinking skills. The final set of research questions investigates how well these instruments can predict attainment in higher education, in terms of the average degree mark obtained by a student. The questions are posed in the light of debates about the selection potential of 'aptitude' tests for higher education in the UK.

The question is posed separately for critical thinking dispositions and critical thinking skills and asks what measures of critical thinking can predict degree results beyond that usually predicted by A-level marks. The analysis was completed using

a hierarchical regression model. The reason for this is that hierarchical regression will identify the additional variance accounted for by each test beyond A-levels, when A-level results have already been included in the model.

Research Question 7

What is the best model of a student's critical thinking skills and dispositions for predicting degree success? (Chapter 7)

Having previously identified significant predictors of degree attainment, a model of critical thinking that best predicts success in higher education is then explored. Predictions are explored for each of the three years of undergraduate study in order to gain an understanding of the developmental relationship between different aspects of critical thinking and examination success that leads to degree outcome in the final year. This analysis is completed with a stepwise regression. A stepwise regression was used because it produces a model of significant predictors; in other words any predictor that does not significantly account for any of the variance is dropped from the model.

4.4 General Design

4.4.1 General Participant Description

All the participants in the research were students at Queen's University, Belfast. Over a period of three years, a total of 387 students took part in the studies, of whom 313 were females and 74 males. Many students participated more than once (the longitudinal sample, see below). Three hundred students had A-level qualifications at entry and 87 had completed gained entry by some other means (e.g., Irish Leaving

Certificate, B-Tech, GNVQ, access courses). Only students with A-level qualifications were used in regression analyses, because of the small numbers in the other groups. All students were studying psychology at some point and were accessed through psychology lectures and 80% graduated with psychology degrees. So generalisations from the sample are restricted.

There were five test periods throughout the course of the research. The first was in April 2001 where participant students completed the CCTDI (N=190) before a first year lab class, at the same time of day, on 1 of 4 weekdays, in the same week.

The second date was in October 2001 when the participant students completed a battery of tests (CCTDI-UK N=119, CCTST N=116, and RAPMsf N=117) during a specially designed lab class for 1st year students. The students attended one of four of these lab classes in the same week.

The third test date was in April 2002 and the participant students completed the CCTDI-UK (N=105) in 2nd year lab classes at the same time of day, on 1 of 4 weekdays, in the same week. These students were a sub-group of students that were tested in the first test period.

The fourth test date was in April 2002 and the participants completed a test battery (CCTDI-UK N=58, CCTST N=43 and RAPMsf N=50) during a single specially designed class for 3rd year students in their module on learning and cognition.

The fifth test session was in April 2003 and the participants completed the CCTDI-UK (N=93) before a single lecture on criminology in their 3rd year. These students were a sub-set of the students tested in the 1st and 3rd test periods.

4.4.2 Cohort Descriptions

Table 4.1 summarises the test period information in section 4.4.1. This table also provides information on the different studies to which each data set is contributed.

Also, sample sizes vary due to incomplete data sets across measures when data were collected in more than one testing session at a particular time point, or when students did not complete their degrees or changed pathways and cannot be tracked.

In order to get maximum use of the data collected, a strategy was adopted to use the greatest number of participants available for each analysis; which results in variable numbers of students in different analyses. Thus, for each study in the thesis, and for each analysis, the sample composition will be carefully described.

Table 4.1 Cohort descriptions detailing study number that some or all of the participants featured in, the tests they completed and the year/time they were tested

Cohort (N)	Test Date	Studies Which used Data	Data Collected (N)	Year/Term of Participants
1 (190)*	April 2001	5.4, 5.5, 5.7, 5.8, 7.2, 7.4	CCTDI(190) CCTDI-UK(190) Degree Marks (146) A-Levels (113)	1/2
2 (119)	Oct 2001	5.6, 5.7, 5.8, 6.4, 6.5, 7.2, 7.3, 7.4	CCTDI-UK (119) RAPMsf (117) CCTST (116) Degree Marks (134) A-Levels (102)	1/1
3 (105)*	April 2002	5.7, 5.8, 6.4, 6.5, 7.4	CCTDI-UK (105) Degree Marks (105) A-Levels (80)	2/2
4 (58)	April 2002	5.7, 5.8, 6.4, 6.5, 7.4	CCTDI-UK (58) RAPMsf (50) CCTST (43) Degree Marks (60) A-Levels (46)	3/2
5 (93)*	April 2003	5.7, 5.8, 7.4	CCTDI-UK (95) Degree Marks (93) A-Levels (67)	3/2

*These participants are sampled from the same group at three consecutive time-points, end of 1st year study, end of 2nd year study and end of 3rd year study. There was some attrition in this longitudinal sample, and some students were tested only once, some were tested twice and some were tested three times. Also, some new students who were not previously tested joined the cohort at each testing point.

Chapter 5

A Psychometric Analysis of Critical Thinking Dispositions: The California Critical Thinking Dispositions Inventory (CCTDI)

5.1 Introduction

The main focus of this chapter is the reliability and validity of critical thinking disposition measurement using the California Critical Thinking Dispositions Inventory (CCTDI). The first section (5.2) is a general discussion on the reliability and validity of psychometric tools. This is followed by more specific information on the psychometric properties of the California Critical Thinking Dispositions Inventory (CCTDI). The remainder of the chapter reports five studies which investigate the psychometrics of critical thinking dispositions testing in undergraduate students. The psychometric techniques used in the chapter include reliability analysis, factor analysis, convergent validity checks and group contrast validity checks.

5.2 The Validity of Psychological Measurement

5.2.1 Introduction

The traditional definition of *validity* is ‘the extent to which a test measures what it is supposed to measure’ (Hogan, 2003, p. 173). However, Hogan (2003) suggests that this definition is not entirely accurate and validity concerns the interpretation of a test score and not the test itself. Therefore, he redefines validity as the appropriate ‘interpretation of a score for a particular purpose or use’ (Hogan, 2003, p. 173).

There are several forms of validity and these different versions have gone through a number of terminology changes since validity classification began. Table 5.1 is taken from Hogan (2003) and helps to clarify these different terminology usages. Then the different types of validity are discussed.

Table 5.1: Outline of the traditional and newer terms for classifying types of validity (adapted from Hogan, 2003)

Traditional Terms	Newer Terms
Content	Content
Criterion Related Concurrent Predictive	Relations to Other Variables (Convergent and Discriminant) Group contrast
Construct	Internal Structure Response Processes Consequences

5.2.1.1 Content Validity (Coverage of domains)

Content validity concerns ‘the relationship between the content of a test and some well-defined domain of knowledge or behaviour’ (Hogan, 2003, p. 177). The main application of content validity is in educational or employment testing. The reason for this is that, when testing for a particular educational domain or employment role, the scope of the domain or role must be covered by the test. Usually the domain is so wide that saturation of all aspects is not feasible; so *sampling* is required to ensure that portions of each aspect are assessed in the test. However, if the sampling is to be adequate, it is important that the domain is well specified either conceptually or pragmatically.

Achieving content validity is a difficult exercise. Firstly, a substantial knowledge of the test domain is required to ensure that all aspects are sampled. For these reasons, a track record of working in the area to be tested, or modelling the test domain by the test author is often given as supporting evidence for content validity. Secondly, the items in the test must be representative of the test domain and this can be difficult for target domains like intelligence that aim to cover a wide range of sub-domains.

5.2.1.2 Criterion Validity (Relations to Other Variables)

Criterion validity deals with the ‘relationship between performance on the test and on some other criterion that is taken as an important indicator of the construct of interest’ (Hogan, 2003, p. 183). Generally, there are two contexts and three applications of criterion validity. The following paragraphs describe those contexts and applications as well some additional considerations.

The two contexts for the use of criterion validity is *predictive* and *concurrent validity*. Predictive validity is based on the premise that a score on a particular test will predict the likelihood of some future behaviour or outcome. An example of this will be seen in Chapter 7 where the predictive validity of scores on the CCTST and the CCTDI are assessed for their ability to predict degree attainment. Concurrent validity is the same in all respects to predictive validity with one difference. Concurrent validity shows relationships between test scores and some other variable of which there is current knowledge. This type of criterion validity will be examined for the CCTDI (5.7) in this chapter and for the CCTST in the next chapter (6.4).

The three main applications of criterion validity are a) external or realistic criterion, b) group contrast validity (group contrasts) and c) convergent validity. An external or realistic criterion is required when the investigator wants to predict what someone might score in a currently unknown measure. Chapter 7 gives an example of this when scores on the CCTST and CCTDI will be used to predict degree attainment. Another reason for using external criterion is that if the actual criterion is too expensive or time consuming to measure and an external criterion test can give a valid estimate of that actual criterion. In both cases the test score is giving a partial estimate of some external criterion.

Another way of demonstrating test score criterion validity is if the test score can differentiate between groups - group contrast validity. An example of group contrast validity will be seen in Section 5.7 when comparisons are made between different year groups of students as they proceed through their undergraduate education. A significant difference between groups in an appropriate direction is usually evidence of this type of criterion validity. The last application of criterion validity is convergent validity. This provides evidence of good criterion validity when a measure shows high correlations with another valid test scores that pertain to measure similar psychological domains (cognitive or affective). Again an example of this will be seen in the thesis when scores on the CCTDI (5.7) and CCTST (6.5) are correlated with the scores of the openness scale on the NEOP-IR, RAPMsf Progressive Matrices, A-levels and Degree marks.

Some additional points need to be considered when discussing criterion validity. The first point is that criterion validity uses correlations and thus it is then affected by the

conditions that affect the correlation coefficient, i.e., linearity, homoscedasticity and homogeneity of the data. The second point is the relationship between reliability and criterion validity. Criterion validity is limited by reliability of the test. Attenuation is the term used to describe the lessening effect imperfect reliability places on validity (Hogan, 2003). A final point worth considering is that high correlations between a test and another test purporting to measure the same construct can be an indicator of criterion validity. Conversely, low correlations between the test and another test measuring an unrelated construct can also provide evidence of validity. These are called convergent and discriminant validity respectively.

5.2.1.3 Construct Validity (Grouping of scores)

Construct validity is the final form of validity discussed in this chapter. In essence, construct validity occurs when certain pieces of evidence suggest that test scores show measurement of a clearly defined construct. The two main techniques for providing this evidence are internal reliability and factor analysis.

Like criterion validity, construct validity is affected by reliability. High internal consistency (generally measured with coefficient alpha) indicates that the test scores are measuring something in a consistent manner and therefore provides evidence that scores are measuring a particular construct. Alternatively, if the internal structure of a test is low as indicated by its consistency level, then the test can be said to have low construct validity. Good internal consistency is essential for construct validity but the test structure can be further improved by factor analysis (Hogan, 2003).

Factor analysis is a powerful tool within psychological measurement for gathering evidence of construct validity. There are two groups of factor analytic methods: confirmatory and exploratory factor analysis. Kline (1998, p. 53) states, ‘Confirmatory factor analysis is used to confirm a hypothesis. For example, the expected factor loadings are put into a target matrix and confirmatory analysis aims to fit this matrix as closely as possible.’ He goes on to say that ‘Exploratory factor analysis was the original use for factor analysis. A large matrix of correlations is factored in order to account mathematically for the correlations in terms of a smaller number of factors. In the terminology of matrix algebra, factor analysis is a procedure for reducing the rank of a matrix, e.g. explaining the correlations between 100 variables in terms of 10 factors.’ In essence, exploratory factor analysis is used to give a statistical representation of which items/variables in a study can be grouped together to form a higher order factor and confirmatory factor analysis is used to check if a factor model is sensible. For more detailed reviews of factor analysis see Cooper (1998) and Kline (1986, 1998 and 2000).

The following Table 5.2 identifies the type of validity assessed in the empirical studies of this thesis.

Table 5.2: Types of validity investigated in the thesis and study numbers in which they feature

Type of Validity	Study Numbers
Construct	1, 2, 3, 5 and 6
Convergent	4 and 7
Group contrast	4 and 7
Predictive	8 and 9

5.3 Previous Psychometric Examination of the CCTDI

The central instrument described in this chapter, the California Critical Thinking Dispositions Inventory (CCTDI), developed in 1992 by Peter Facione, was the first test to explicitly measure critical thinking dispositions (see section 3.3.6). Content validity is a strong point of the test, because it originated from the Delphi Report (1990) where an expert panel mapped out the conceptual domain covered by the term critical thinking. Nevertheless, the factor structure of the CCTDI, i.e., the test sub-scales, have dubious validity, particularly in relation to its criterion and construct validity. There are a number of reasons for suggesting that the CCTDI has poor validity. The first reason is that the CCTDI manual (2000) contains little evidence of factor analytic support for the structure of the CCTDI sub-scales. Secondly, reviews by Callahan (1995) and Ochoa (1995) featured in the mental measurement yearbooks (Conoley and Impara, 1995) were unsupportive of the factor structure of this measurement tool. Thirdly, as the inventory is being used for the first time in a UK context it is unclear what the cultural influences on the factor structure might be.

Two previous attempts to improve the sub-scale validity of the CCTDI have been reported, initially by Walsh and Hardy (1997) with a cross-validity study and then by Kakai (2003) which added to the previous work. Walsh and Hardy examined the factor structure of the CCTDI and compared the sub-scales validity across gender and various students' degree disciplines in the US (N=499). They found that the test was 'highly stable' across genders for the Truth-seeking and Open-mindedness subscales; the Systematicity and Confidence sub-scales were 'moderately stable'; and the Analyticity, Maturity (Facione et al, 1992) and Inquisitiveness subscales were not stable. They also found that only the truth seeking factor structure was

stable across practice (Nursing, Education and Business) and non-practise degrees (English, History and Psychology). The analysts suggested that further refinement of the scale was necessary. Kakai's (2003) factor analytic study of a sample of Hawaiian students (N=536) responses on the CCTDI suggested that there were four factors in the test. He also suggested that further refinement was necessary and that a shorter reconfigured form of the test may be of experimental use.

In the context of some doubt about the stability of the factor structure of the CCTDI, the following studies were conducted. The general strategy adopted was to attempt to maintain the content validity of the inventory but to improve the sub-scale reliability/validity by shortening the test, re-arranging items and so on, with the ultimate purpose of producing a psychometrically sounder version for use within UK populations.

5.4 Study 1 Internal Consistency and Confirmatory Factor Analysis of the CCTDI

5.4.1 Introduction

The purpose of this study is to check on the reliability and construct validity of the CCTDI for a UK undergraduate population. The inventory claims to measure a range of critical thinking dispositions derived from the Delphi analysis of the meaning of critical thinking. There are seven subscales; the following paragraphs name each sub-scale, give a short description of its meaning taken from the test manual and an example item from the test (Facione et al., 2000).

Inquisitiveness (I – scale, 10 items) ‘Measures one’s level of intellectual curiosity.’ (Facione et al. 2000, p.3).

‘Studying new things all my life would be wonderful’

Critical Thinking Self-Confidence (C – scale, 9 items) ‘Refers to the level of trust one places in one’s own reasoning processes.’ (Facione et al., 2000, p. 3)

‘I pride myself on coming up with creative alternatives.’

Truth-Seeking (T – scale, 12 items) ‘Targets the disposition of being eager to seek the truth, courageous about asking questions, and honest and objective about pursuing inquiry even if the findings do not support one’s interests or one’s preconceived opinions.’ (Facione et al., 2000, p. 2)

‘It’s impossible to know what standards to apply to most questions.’

(Negatively Marked)

Open-Mindedness (O – Scale, 12 items) ‘Targets the disposition of being open-minded and tolerant of divergent views with sensitivity to the possibility of one’s own bias.’ (Facione et al., 2000, p. 2)

‘Others are entitled to their opinions, but I don’t need to hear them.’
(Negatively Marked)

Analyticity (A – Scale, 11 items) ‘Targets the disposition of being alert to potentially problematic situations, anticipating possible results or consequences, and prizing the application of reason and the use of evidence even if the problem at hand turns out to be challenging or difficult.’ (Facione et al., 2000, p. 3)

‘When I have to deal with something really complex it’s panic time.’
(Negatively Marked)

Systematicity (S – Scale, 11 items) ‘Targets the disposition toward organized, orderly, focused and diligent inquiry.’ (Facione et al., 2000, p. 3)

‘When faced with a big decision, I first seek all the information I can.’

Maturity (M – Scale, 10 items) ‘How disposed a person is to make reflective judgements.’ (Facione et al., 2000, p. 3)

‘I believe what I want to believe.’
(Negatively Marked)

5.4.2 Method

Participants: Study 1 had 190 participant students (see Table 4.1, Cohort 1) who were in their first year of university study. There were 52 males and 138 females in Cohort 1. Over two-thirds of the students in this cohort were intending to continue studying psychology for their degree and the remaining one-third of the students were studying psychology in first year but were registered on other degree pathways. All students were attending a first year laboratory class called Basic Psychology 2 in their second semester. The test dates were from 4/4/2001 to 7/4/2001.

Table 5.3: Breakdown of degree pathways for the students in Cohort 1

Degree Major	Frequency	Percent
Psychology	125	65.8
Computers	25	13.1
Geography	16	8.4
Sociology	7	3.7
Bio Science	5	3.1
Others	12	5.9
Total	190	100.0

Procedure: The inventory was administered as part of the students' regular laboratory sessions. Students completed the test within 20 minutes but no time limit was imposed. The test was administered as per the instructions in the CCTDI manual (Facione et al., 2000, p. 9). The investigator explained to the students that their names were requested on the test paper for administrative purposes, as they may be asked to complete further questionnaires related to the study at some future time and names would facilitate tracking the data across time points. However, they were assured that all their responses would be treated confidentially, that responses would

not be revealed to course tutors and that only aggregated data, not individual data, would be reported. It was indicated that students could withdraw from the study at any time.

Data Analysis: There are two parts to the data analysis for the study; the first analysis examines the internal consistency of the tests' subscales and this was investigated using Cronbach's alpha. The second part examined the construct validity of the CCTDI subscale model. This was examined using confirmatory factor analysis on each set of subscale scores.

5.4.3 Results

Table 5.4 shows the Cronbach's alphas for each of the 7 sub-scales of the CCTDI. Alpha is a measure of internal consistency, i.e., how consistent the scores are on the items within the sub-scale. For example, if a student scores highly on an item on the inquisitiveness scale, then s/he should be consistently scoring highly on the other inquisitiveness items.

Table 5.4 shows the reliability of the sub-scales in descending order. For example the Self-confidence sub-scale is the most reliable with an alpha of .78 and the most inconsistent scale is the open-mindedness scale with an alpha of .50. Overall the reliability analysis shows that only two out of the seven sub-scales had a reliability above the .7 threshold (CT Self-Confidence and Inquisitiveness), which is generally considered acceptable for a good psychometric test (Cooper, 1998).

Table 5.4: Internal consistency of the CCTDI sub-scales and the number of items in each sub-scale

Sub-scale	Alpha (N=190)	No. of Items
CT Self-Confidence	.78	9
Inquisitiveness	.72	10
Analyticity	.65	11
Truth Seeking	.60	12
Systematicity	.59	11
Maturity	.56	10
Open-Mindedness	.50	12

Table 5.5 shows the results of a confirmatory factor analysis. The technique used was a maximum likelihood factor analysis with 1 factor extracted. Maximum likelihood was used as only 1 factor was being extracted in each of the 7 analyses (max = 12). The reason for this is that maximum likelihood analyses does not assume that the extracted factor or factors account for all of the variance in the construct. In this case it was assumed that extracting one factor would not account for all of the variance in each construct. A chi-square measure is used to test the goodness of fit of the data against the expected model. Contrary to most types of statistical analyses, the absence of statistical significance is preferable because it shows less difference between the expected structure for a factor and the observed factor structure. In other words the chi-square tests the difference between the perfect sub-scale model and the empirical data.

The results of Table 5.5 shows that there was a significant difference between the observed and expected sub-scale model in each of the 7 cases. Therefore the factors were not confirmed in any instance.

Table 5.5: Confirmatory factor analysis of the CCTDI sub-scales

Sub-scale	Goodness of Fit Test		
	Chi-Square	df	Sig.
Maturity	55.7	35	.015
Inquisitiveness	57.9	35	.009
Systematicity	69.4	44	.009
CT Self-Confidence	54.6	27	.001
Analyticity	86.0	44	.000
Open-Mindedness	100.7	54	.000
Truth-Seeking	125.2	54	.000

5.4.4 Conclusions

The reliability and factor analysis in this study concurs with the evidence provided by the other investigators (Walsh and Hardy, 1997; Kakai, 2003) mentioned in the introduction to this chapter. In short, this scale has poor construct reliability and validity – at least for this population. Therefore there are insufficient grounds to proceed with the inventory in its published form. The next step then was to proceed with an exploratory factor analysis with the intention of improving reliability and construct validity.

5.5 Study 2: Exploratory Principle Component and Reliability Analysis of the CCTDI

5.5.1 Introduction

The previous study provided additional evidence of poor construct validity and reliability of the CCTDI. This study attempts to improve the reliability and validity of the CCTDI by restructuring and eliminating items from the test. Firstly, an exploratory principal component analysis was carried out on the test scores to identify the number of factors and their constituent items thus providing construct validity. These ‘new’ factors were then assessed for reliability with a Cronbach’s Alpha test of internal consistency.

5.5.2 Method

Participants: As this is essentially a second analysis of the data set collected for Study 1, the participants are identical to the ones in the previous study.

Procedure: As in Study 1.

Data Analysis: The analysis technique used here was a principal components analysis which is the simplest form of factor analysis. There was a large number (75) of items in the CCTDI which can be assumed to account for a large proportion of the variance in critical thinking dispositions; this is consistent with the assumptions of principle component analysis that the factors account for all the variation between the variables. Regardless of the proportion of variance accounted for by the analysis, principle component analysis will provide highly similar solutions to any form of factor analysis (Tabatchnick and Fidell, 1989; Kline 1994). The analyses were carried out as follows: (1) the initial number of factors was identified through examining the scree plot; (2) factors were then rotated using Varimax rotation; (3)

factor loadings for the items were examined and the names for factors were decided upon. Some items did not load sufficiently on any factor and were subsequently dropped from the inventory; other items loaded on more than one factor and decisions had to be made about where to place the item. In addition, issues about the wording and semantic interpretation of items that may have contributed to alternative interpretations were dealt with; (4) then the reliability of the 'new' scales were checked.

5.5.3 Results - Study 2(a): Exploratory Factor Analysis of the CCTDI

Part 1 Factor extraction and rotation: A principal components analysis was carried out on the 75 items of the CCTDI. Both the Kaiser-Meyer-Olkin Measure of sampling adequacy (.653) and Bartlett's test of sphericity ($X^2 = 5493$, $df = 2775$, $p < .000$) showed suitability of the data for component analysis. The scree plot (see Figure 5.1) suggests that there were 7 factors. Therefore 7 factors were extracted, which explained 35.8% of the variance. An orthogonal (uncorrelated) Varimax rotation was carried out on the 7 factors as the factor correlation matrix suggests there was no single higher order factor because of the number of low and non-significant correlations (see component transformation matrix Table 5.6). The rotation converged in 15 iterations.

Figure 5.1: Scree plot for the exploratory factor analysis on the responses of Cohort 1 on the 75 items of the CCTDI

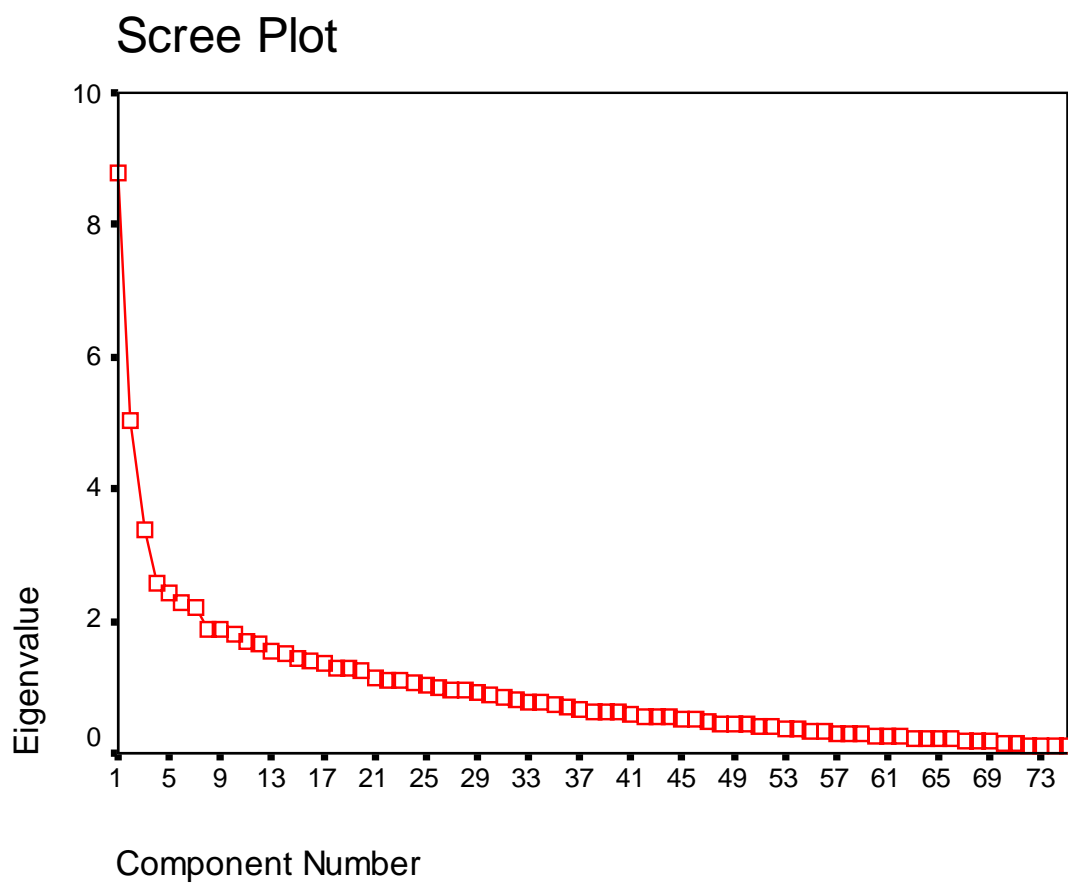


Table 5.5 This Component Transformation Matrix shows that there are low correlations between the factors in the CCTDI; thus it is not advised that these subscales be added together to form an overall score for critical thinking dispositions.

Key C=Self-Confidence, O=Open-mindedness, A=Analyticity, S=Systematicity, T=Truth-seeking, I=Inquisitiveness and M=Maturity.

Table 5.6: Component correlation matrix for the seven CCTDI subscales

Component	A	T	C	O	S	M	I
A	1.000	.171	.084	.027	.135	-.067	.111
T	.171	1.000	.026	.130	.027	-.115	.112
C	.084	.026	1.000	-.035	.187	.085	.145
O	.027	.130	-.035	1.000	-.033	-.076	.097
S	.135	.027	.187	-.033	1.000	.024	.030
M	-.067	-.115	.085	-.076	.024	1.000	-.040
I	.111	.112	.145	.097	.030	-.040	1.000

Examining the Factor Loadings from the Exploratory Factor Analysis:

(N.B the factor loadings for all items are included in Appendix 5.1)

Several points emerged from the initial examination of the factor loadings of items. Twenty two items (from a total of 75) did not reach the guideline cut-off point of .40 or above. Thirteen items that did not reach the numerical cut-off point were dropped and nine were kept because they were close to the cut off point (factor loadings that were between 3.4 and 3.9) and were considered suitable for the factor. Also, eleven items loaded highly on more than one factor, indicating that the emergent factor structure might not be clear. Three of the items loading highly on two factors were removed because the questions were ‘double barreled’ in nature, i.e., they were referring to two meanings and could be ambiguously interpreted. Eight other items

loaded on more than one factor but were retained. These items were placed in a single factor using one or both of the following guidelines: if the loadings were clearly in favour of a single factor, then the item was included on the factor with the highest loading or if the item when closely examined fitted a particular factor then it was placed in the factor of best fit. Eighteen items in total were removed (see items in bold in Tables 5.7-5.13 for specific reasons for omission) thus leaving 57 items from the original 75 on the CCTDI. One item was altered (CCTDI Item 51) and added to the 57 retained items. However, one item was omitted due to a clerical error (CCTDI Item 38) thus leaving 57 items in total. This item loaded at .43 on the Systematicity sub-scale. Tables 5.7 – 5.13 show the factor loadings for the final 57 items on all seven factors.

Guidelines for interpreting Tables 5.6 - 5.12

Factor Loading CCTDI-UK: The items in each sub-scale are rank ordered (highest first by its factor loading on the revised sub-scale). An asterisk de-notes that the item loads highly on more than 1 factor above .34. The numbers in the factor loading column shows those factors on which the item loads highly. The underlined number denotes the factor on which the item was eventually placed.

New No. on CCTDI-UK: This shows the item numbering on the revised inventory.

CCTDI sub-scale: This names the sub-scale to which the item belonged on the original CCTDI. It is apparent that there are clusters of these items in the new sub-scale tables. The same key is used as in Table 5.6.

Old No. on CCTDI: These are the item numberings for the original CCTDI.

Question and reason for deletion if applicable: This column shows the wording of each item. The items in **bold writing** have been removed from the new version of the test and the reason why is given in *italics* underneath the item.

Factor 1 Critical Thinking Self Confidence

Factor 1 had sixteen items loaded positively above .40 and six of them loaded on more than one factor. Nine of the items were from the Self-Confidence sub-scale of the original inventory and the remainder were from three other subscales. The items that loaded most highly were strong on self-evaluation about the person's own ability or how others might view them. Examples of the highest loading items were 'being inquisitive is one of my strong points' 'others admire my intellectual curiosity and inquisitiveness', 'others look to me to establish reasonable standards to apply to decisions'. Hence it was decided to call this factor Critical Thinking Self-Confidence, as it resembled most closely the original self-confidence subscale. Three items that loaded on more than one factor were retained on Factor 1 because of this 'others' perspective. Two items were subsequently dropped because of cross-factor loading and/or ambiguous interpretation. Fourteen items remained to be included in the new sub-scale.

Table 5.7: Factor 1 Critical Thinking Self-Confidence

Load on -UK	New No.	CCTDI Scale	Old No.	Question and Reason for removal if applicable
.656	29	C	40	I pride myself on coming up with creative alternatives.
.649	37	C	49	Being inquisitive is one of my strong points.
.581	35	C	46	Others look to me to establish reasonable standards to apply to decisions.
.564	15	C	18	Others admire my intellectual curiosity and inquisitiveness.
.557	24	O	30	It's important to me to understand what other people think about things.
.531	31	A	42	Frequently I find myself evaluating other people's arguments.
.522* 1,7	40	C	52	I take pride in my ability to understand the opinions of others.
.502	22	C	27	My peers call on to make judgements because I decide things fairly.
.501	5	A	6	It bothers me when people rely on weak arguments to defend good ideas
.489* 1,3	52	A	69	Others look to me to decide when the problem is solved. Loads on 1 (.489) and 3 (.454) put in 1 again b/c of "others", which pertains towards to self image.
.444* 1,5		I	55	I really enjoy trying to figure out how things work. <i>Taken out because it loads on two factors; 1 (.444), and 5 (.423). Also it has double barrelled nature.</i>
.439* 1,3		A	54	You could describe me as logical. <i>Taken out because loads on two factors, 3 (.538) and 1 (.439). Also it is ambiguous because of the word "could."</i>
.436	8	C	10	I'm proud that I can think with great precision.
.423* 1,3	42	C	56	Others look to me to keep working on a problem when the going gets tough. Loads on 1 (.423) and 3(.439). Left in 1 because of "others" thus displaying self-confidence.
.404*		I	47	I look forward to learning challenging things.
.40	13	C	16	Tests that require thinking, not just memorization, are better for me.
.373	6	O	8	It concerns me that I might have biases of which I am not aware. <i>Kept in because it suited factor (self perception).</i>

Factor 2 Truth Seeking

Factor 2 had nine items which loaded positively above .40. There was no cross loading in this factor. Three of the items were from the Truth-Seeking sub-scale (including the two highest loading items) of the original inventory and the remainder were from four other subscales. All the items on this sub-scale are reverse marked items, i.e., an overall high score reflects a low level in this disposition. The items that loaded most highly refer to answering questions about the relativity and worth of different opinions and evaluating points of view. Examples of the highest loading items were 'It is impossible to know what standards to apply to most questions', 'It is never easy to decide between competing points of view'. Hence it was decided to call this factor Truth-Seeking, as it resembled most closely the original truth-seeking scale and there was an overlap in items. Nine items remained to be included in the new sub-scale.

Table 5.8: Factor 2 Truth-Seeking

Load on -UK	New No.	CCTDI Scale	Old No.	Question and Reason for removal if applicable
.580	54	T	72	It's impossible to know what standards to apply to most questions.
.568	4	T	5	It's never easy to decide between competing points of view.
.507	1	O	1	Considering all the alternatives is a luxury I can't afford.
.481	46	A	60	There is no way to know whether one solution is better than another.
.478	44	S	58	My opinion on controversial topics depends a lot on who I talk to last.
.474	41	M	53	Analogies are about as useful as a sailboat on a freeway.
.474	10	T	12	If there are four reasons in favour and one against, I'd go with the four.
.450	34	O	45	I shouldn't be forced to defend my own opinions.
.416	11	M	14	Advice is worth exactly what you pay for it.
.372		M	32	Reading is something I avoid, if possible. <i>Doesn't load above 0.4 on any factor. It also loads .362 on 5.</i>
.306		M	3	The best argument for an idea is how you feel about it at the moment. <i>Doesn't load above 0.4 on any factor.</i>

Factor 3 Analyticity

Factor 3 had eleven items loading positively above .40 and three loaded close enough to .4 to be considered. Four items loaded on more than one factor. Five items were from the Analyticity sub-scale of the original inventory and the remainder were from five other subscales. The items that loaded most highly refer to fear of complexity and using reluctance to use logical reasoning. Examples of the highest loading items were ‘When I have to deal with something really complex it’s panic time.’ ‘I pretend to be logical but I’m not.’ Hence it was decided to call this factor Analyticity, as it resembled most closely the Analyticity subscale. New versions of three items were retained in this sub-scale to reduce confusion and reverse polarity to fit with the other items. One of these items loaded on more than one factor; so a new version of this item was retained on Factor 3. Two items that loaded on other factors were moved because they fitted better with other factors and/or loaded more highly on the alternative factor. Three items were subsequently dropped because of ambiguous interpretation and low factor loading. Nine items remained to be included in the new sub-scale.

Table 5.9: Factor 3 Analyticity

Load on -UK	New No.	CCTDI Scale	Old No.	Question and Reason for removal if applicable
.689	26	T	35	When I have to deal with something really complex it's panic time.
.601	18	A	21	I pretend to be logical but I'm not.
.551	28	S	37	People think I put off making decisions. 'Procrastinate' changed to "put off" in order to reduce confusion.
.538* 1,3		A	54	You could describe me as logical. <i>Taken out because loads on two factors, 3 (-.589) and 1 (.439). Also it is ambiguous 'could.'</i>
.529	50	A	66	Life has taught me not to be too logical.
.514		A	63	I'm known for approaching complex problems in an orderly way. Taken out because of polarity and difficult sentence structure.
.454* 1,3		A	69	Others look to me to decide when the problem is solved. Loads on 1 (.489) and 3 (.454) put in 1 again b/c of "others", which pertains towards to self image.
.444* 3,7		S	74	I'm good at developing orderly plans to address complex problems. <i>Placed in 7 because it fits better and 7 needed more items.</i>
.439* 1,3		C	56	Others look to me to keep working on a problem when the going gets tough. Loads on 1 (.423) and 3(.439). Left in 1 because of "others" thus displaying self-confidence.
.426	47	M	61	The best way to solve problems is to ask somebody else for the answers.
.411	3	S	4	My trouble is that I'm easily distracted.
.390	14	S	17	I can talk about my problems for hours without solving anything. Left in because it was close to .4 and seemed to fit factor well
.389	19	S	22	It's hard for me to organise my thoughts. 'it's easy' was changed to "it's hard" to load in a consistent direction and left in because it loaded quite close to .4.
.359		T	62	Many questions are just too frightening to ask. Doesn't load above 0.4 on any factor.
.352* 3,5	58	I	51	Complex problems are fun to figure out. Loads on 3 (.352) and 5 (.468). Left in 5 because it fits factor better. Another version of this question was produced with 'hard' instead of 'fun' and placed in this factor.

Factor 4 Open-mindedness

Factor 4 had six items which loaded positively above .40 and two items loaded close enough to be considered. One item loaded on more than one factor. Five items were from the Open-mindedness sub-scale of the original inventory and the remainder were from five other subscales. All the items in this factor were reverse marked items. The items that loaded most highly show a lack of tolerance for alternative opinions. Examples of the highest loading items were ‘Others are entitled to their opinions, but I don’t need to hear them.’ ‘Things are, as they appear to be.’ ‘Being open-minded about different worldviews is less important than people think’. Hence it was decided to call this factor Open-mindedness, as it had substantial overlap with the original open-mindedness scale and included a number of the original open-mindedness items. The one item that cross-loaded was placed in the alternate factor as it loaded higher on that factor. Five items were subsequently dropped because they didn’t load above .4. Seven items remained to be included in the new sub-scale.

Table 5.10: Factor 4 Open-Mindedness

Load on -UK	New No.	CCTDI Scale	Old No.	Question and Reason for removal if applicable
.571	55	O	73	Others are entitled to their opinions, but I don't need to hear them.
.450	51	M	67	Things are, as they appear to be.
.443	48	O	64	Being open-minded about different worldviews is less important than people think.
.443	12	S	15	Most college courses are uninteresting and not worth taking.
.435	17	O	20	You are not entitled to your opinion if you are obviously mistaken.
.40* <u>6,4</u>		T	75	To get people to agree with me I would give any reason that worked. Loads on 4 (.40) and 6 (.410) put in 6 because it fits factor better and loads marginally higher.
.383	23	M	28	Being open-minded means you don't know what's true and what's not. <i>Doesn't load above .4 but kept in because it suited factor.</i>
.382	27	O	36	Foreigners should study our culture instead of us always trying to understand theirs. <i>Doesn't load above .4 but kept in because it suited factor.</i>
.379		M	71	Powerful People Determine the right answer. <i>Doesn't load above 0.4 on any factor.</i>
.371		I	34	Required subjects in college waste time. <i>Doesn't load above 0.4 on any factor.</i>
-.358		S	68	If I have to work on a problem I can put other things out of my mind. <i>Doesn't load above 0.4 on any factor.</i>
.337		M	7	The truth always depends on your point of view. <i>Doesn't load above 0.4 on any factor.</i>
.275		O	48	It makes a lot of sense to study what foreigners think. <i>Doesn't load above 0.4 on any factor.</i>

Factor 5 Inquisitiveness

Factor 5 had eight items which loaded positively above .40 and one item loaded close enough to be considered. There were no cross loading items in this sub-scale. 7 out of the eight eventual items in this scale were from the original inquisitiveness sub-scale. The one other item was from the truth-seeking sub-scale. The items that loaded most highly show an enjoyment of discovery and learning. Examples of the highest loading items were ‘Studying new things all my life would be wonderful.’ ‘I look forward to learning challenging things’ ‘I really enjoy trying to figure out how things work.’ This factor was called inquisitiveness because of the overlap with the original sub-scale and the themes running through the top loading items. Two items were slightly altered to fit the factor better. Two items were subsequently dropped because they didn’t load above .4 and one item was removed because it didn’t suit the factor. Seven items remained to be included in the new sub-scale.

Table 5.11: Factor 5 Inquisitiveness

Load on -UK	New No.	CCTDI Scale	Old No.	Question and Reason for removal if applicable
.497	2	I	2	Studying new things all my life would be wonderful.
.457	36	I	47	I look forward to learning challenging things.
.457	45	I	59	No mater what the topic, I am eager to know more about it.
.446		I	55	I really enjoy trying to figure out how things work.
.446		O	24	Open-mindedness has limits when it comes to right and wrong. <i>Taken out because it didn't fit factor well.</i>
.439	39	I	51	Complex problems are fun to figure out.
.432	49	I	65	Learn everything you can you never know when it might come in handy.
.428	33	I	44	It's <u>really</u> important to keep trying to solve difficult problems. 'Just not that' changed to 'Really' in order to fit factor.
.368		M	32	Reading is something I avoid, if possible. <i>Taken out because it doesn't load above .4 on any factor.</i>
.368	38	T	50	<u>I enjoy collecting information</u> to support my views. Question kept in because it fits factor also changed from 'I look for facts that support my views, not facts that disagree,' to fit factor better i.e. reduce its double barrelled nature.
.309		O	13	Men and women are equally logical. <i>Doesn't load above 0.4 on any factor.</i>

Factor 6 Maturity

Factor 6 had five items loading positively above .40. All the items in this sub-scale are reverse marked items. There were no cross loading items in this sub-scale. One item in this scale was from the original maturity sub-scale. The other four selected items were from the truth-seeking sub-scale. The items that loaded most highly show rigidity and staunchness of beliefs. Examples of the highest loading items were ‘I believe what I want to believe.’ ‘Even if the evidence is against me, I’ll hold firm to my beliefs’ ‘I know what I think, so why should I pretend to ponder my choices.’ This factor was called maturity because of the themes running through the top loading items. Three items were subsequently dropped because they didn’t load above .4. Five items remained to be included in the new sub-scale.

Table 5.12: Factor 6 Maturity

Load on -UK	New No.	CCTDI Scale	Old No.	Question and Reason for removal if applicable
.587	32	T	43	I believe what I want to believe.
.536	16	T	19	Even if the evidence is against me, I’ll hold firm to my beliefs.
.533	53	T	70	I know what I think, so why should I pretend to ponder my choices.
.410	57	T	75	To get people to agree with me I would give any reason that worked.
.40	9	M	11	We can never really measure the truth about most things.
.365		S	33	People say I rush into decisions too quickly. Doesn’t load above 0.4 on any factor.
-.351		O	41	Frankly, I am trying to be less judgemental. Taken out because it doesn’t load on any factor above .40.
-.327		T	23	Everyone always argues from their own self-interest even me. Doesn’t load above 0.4 on any factor.

Factor 7 Systematicity

Four items loaded positively above .40 on Factor 7 and two items loaded close enough to .4 to be considered. There was one cross loading item in this sub-scale which was retained in this sub-scale because of its appropriateness and requirement for items in this scale. Three items in this scale were from the original Systematicity sub-scale. The other three retained items were from the Analyticity and Inquisitiveness sub-scales. The items that loaded most highly show organisation in thinking and focus on the problem. Examples of the highest loading items were ‘When faced with a big decision, I first seek all the information I can.’ ‘Getting a clear idea about the problem at hand is the first priority’. This factor was called Systematicity because of the overlap in items with the original sub-scale and the theme running through the top loading items. Six items remained to be included in the new sub-scale.

Table 5.13: Factor 7 Systematicity

5.5.4 Results - Study 2(b): Reliability Analysis of the CCTDI-UK

Load on -UK	New No.	CCTDI Scale	Old No.	Question and Reason for removal if applicable
.580	21	I	26	When faced with a big decision, I first seek all the information I can.
.550	43	A	57	Getting a clear idea about the problem at hand is the first priority.
.497	20	S	25	It's important to me to keep careful records of my personal finances.
.454	7	S	9	I always focus the question before I attempt to answer it.
.375	25	A	31	I must have grounds for all my beliefs. <i>Kept in because it fits factor.</i>
.344* 3, <u>7</u>	56	S	74	I'm good at developing orderly plans to address complex problems. Loads on 3 (.444) and 7 (.344) put in 7 because it fits better with a focus on being systematic.

Table 5.14 shows a reliability analysis of the newly derived sub-scales as described in Table 5.7-5.13. Reducing item numbers and improving reliability is a recognised method of increasing construct integrity. Items were removed from the sub-scales if their removal increased the sub-scale reliability. There was a total of 12 items removed as a result of the reliability analysis. Five of the seven scales had items removed, which resulted in an improvement in their reliability in all five cases. The maturity sub-scale was removed from further studies because it was left with only two items and had a relatively low reliability.

Table 5.14: Internal Consistency (Cronbach's Alpha) of revised sub-scales and internal consistency when low reliability items were removed

Sub-Scale	Items	Alpha	Items Removed	New Alpha	Number of items remaining
CT Self-Confidence	5,6,8,13,15,22,24,29,31,35,37,40,42,52	.821	5,6,13	.848	11
Inquisitiveness	2,33,36,38,39,45,49	.814	None	.814	7
Analyticity	3,14,18,19,26,28,47,50,58	.696	50,58	.736	7
Systematicity	7,20,21,25,43,56	.665	None	.665	6
Truth Seeking	1,4,10,11,34,41,44,46,54	.601	1,10	.625	7
Open-Minded	12,17,23,27,48,51,55	.560	17,51	.579	5
Maturity	9,16,32,53,57	.362	9,16,57	.517	2

5.5.5 Conclusions

An exploratory factor analysis revealed a seven factor solution from the data. It was agreed that the item arrangement and themes in the solution model sufficiently mapped onto the original CCTDI sub-scales to retain the original seven scale names. However, to avoid confusion between the original inventory and the restructured scale, it was renamed the CCTDI-UK because it was devised using a UK population. The main reasons for removing items from the test were low loading and cross loading items. The cross loading items often had a double barrelled nature in that they were making two points in the one statement.

The reliability analysis further refined the sub-scales of the CCTDI-UK by the removal of unreliable items. The maturity sub-scale was found to psychometrically unstable and thus was removed from further analysis.

5.6 Study 3: Confirmatory Factor Analysis of the CCTDI-UK and Reliability construct validity comparison with the CCTDI

5.6.1 Introduction

The last two studies improved the reliability and construct validity of the CCTDI to produce the CCTDI-UK. The current study examines the construct validity of the CCTDI-UK with a confirmatory factor analysis on a new sample of undergraduate students. Following that the two versions of disposition inventory are compared for reliability and construct validity.

5.6.2 Method

Participants: The third study was a confirmatory factor analysis of the CCTDI-UK. The sample used here were 119 undergraduate psychology students (see section 4.3.2 Cohort 2) at Queen’s University, Belfast in their first year of study. There were 106 females and 13 males in the sample.

Data analysis: The basic design is the same as in study the confirmatory factor analysis in Study 1. Furthermore there is a comparison of reliability and construct validity between the CCTDI and CCTDI-UK.

5.6.3 Results Study 3(a) - Confirmatory Factor Analysis of the CCTDI-UK and comparison with the CCTDI

Table 5.15 is a replication of the confirmatory factor analysis carried out in Section 5.4 using the revised CCTDI-UK subscales and a new population of participants. A maximum likelihood factor analysis with one factor extracted for the reasons outlined in 5.4. The significance tests from the CCTDI confirmatory analysis are included in the table for ease of comparison. Comparing the CCTDI with the CCTDI-UK it can be seen that the significance levels increased in five out of six cases, i.e., there is an improvement in the construct stability in these five cases. Two of the six sub-scales (Open-minded and Systematic) now show no difference between their observed structure and the expected factor structure. The Truth-seeking sub-scale is approaching non-significance. This compares favourably to the original version where all sub-scales showed a significant difference between the observed and expected model.

Table 5.15: Confirmatory Factor Analysis of the CCTDI-UK sub-scales. CCTDI confirmatory analysis significance levels are included

	CCTDI-UK Goodness of Fit Test			CCTDI Goodness of Fit Test significance
Sub-scale	Chi-Square	df	Sig.	Sig.
Open-Minded	6.5	5	.260	.000
Systematic	15.2	9	.086	.009
Truth-seeking	24.3	14	.042	.000
CT Self-Confidence	69.2	44	.009	.001
Analyticity	35.4	14	.001	.000
Inquisitive	39.5	14	.000	.009
Maturity	removed			

5.6.4 Study 3(b) - Reliability Comparison between the CCTDI-UK and the CCTDI

Table 5.16 shows no additional analysis of the CCTDI-UK but places the reliability analysis of the CCTDI-UK (Table 5.14) and CCTDI (Table 5.4) together for ease of comparison. The alphas of all scales in the CCTDI-UK show an improvement over the CCTDI.

Table 5.16: Internal Consistency and number of items for each sub-scale of the CCTDI-UK and CCTDI

Sub-Scale	CCTDI-UK Alpha	Number of items CCTDI-UK	CCTDI Alpha	No. of Items CCTDI
CT Self-Confidence	.85	11	.78	9
Inquisitiveness	.81	7	.72	10
Analyticity	.74	7	.65	11
Systematicity	.67	6	.59	11
Truth Seeking	.63	7	.60	12
Open-Minded	.58	5	.50	12
Maturity	Removed			

5.6.5 Conclusions

There was an improvement in construct validity in 5 out of the 6 revised sub-scales. In fact three out of the six revised sub-scales non-significant or approaching a non-significant difference between the observed and expected model (evidence of good construct validity), whereas, none of the CCTDI sub-scales were close to being non-significant. There was a reduction in the number of items from the CCTDI to the CCTDI-UK in 5 of the six sub-scales, yet there was still an increase in the reliability of all six sub-scales. Reducing items will normally reduce reliability (Cooper, 1998); so the observed increase in reliability is further evidence of sub-scale consistency. In short, there was an improvement psychometric strength of the CCTDI-UK over the CCTDI.

5.7 Study 4: External Criterion Validity (Convergent and Group Contrast Validity) of the CCTDI-UK

The previous studies in this chapter investigated the internal structure of the CCTDI-UK, i.e., construct validity. These studies looked at internal consistency and factor structure which are both indicators of construct validity. The following analysis investigates two new forms of criterion validity of the CCTDI-UK, namely convergent validity and group contrast validity. These types of validity are investigated using external criteria, hence the name criterion validity.

5.7.1 Study 4(a): Convergent Validity of CCTDI-UK

5.7.1.1 Introduction

Correlation with another test which measures a similar construct is a further way of investigating validity. The CCTDI-UK sub-scales are inter-correlated because they are theoretically measures of a related construct, i.e. critical thinking disposition. These scales were also correlated with a theoretically similar domain, Openness, from the Revised NEO Personality Inventory (NEO PI-R). The NEO PI-R is a widely used measure of the five major dimensions of personality Neuroticism, Extraversion, Openness, Agreeableness and Conscientiousness. The facets within the Openness sub-scale were considered to have some overlap with the CCTDI sub-scales as they included constructs such as Ideas, and Actions. Furthermore, the following sample items from the NEO PI-R should demonstrate the suitability of the scale.

- I think it is interesting to learn and develop new hobbies.
- I often enjoy playing with theories or abstract ideas.
- I enjoy solving problems or puzzles. (Costa and McCrae, 1991, p.71)

The NEO PI-R is a well researched scale and has been shown to have strong internal and external validity. There is comprehensive reliability and validity (both construct and criterion) information contained within the NEO PI-R manual (Costa and McCrae, 1991, p. 39 - 55).

5.7.1.2 Method

Participants: There were a total of 372 participants in this study who had completed the CCTDI-UK. However, only 95 of these students had completed both the CCTDI-UK and the NEO PI-R Openness sub-scale, thus accounting for the different participant numbers in the analysis in Table 5.18. The 372 participants were drawn from all three years of study in a higher education degree and included some students from all cohorts 1-5. However longitudinal students' data were used from one of the three time points. Their numbers and order of selection preference are broken down in the following table.

Table 5.17: The number of participants from each of the cohorts who participated in the study

Cohort (X)	Number of participants
Exit (4)	58
Entry (2)	119
Longitudinal 3 (5)	95
Longitudinal 1 (1)	85
Longitudinal 2 (3)	15

Statistical Analysis: The study was a convergent validity investigation of the CCTDI-UK, which inter-correlated the sub-scales of the CCTDI-UK (N=372) and a sub-set of these students' scores were correlated with the openness sub-scale of the NEO-PIR. However only a sub-sample completed this questionnaire (N=95).

5.7.1.3 Results

Table 5.18 below shows the correlations between the sub-scales of the CCTDI-UK and the Openness subscale from the NEOP-IR. The first point to note about the inter-correlations between the sub-scales of the CCTDI-UK is that 13 out of 15 possible correlations are statistically significant. The sub-scale, Systematicity has almost zero correlation with Open-mindedness and Truth-seeking, but does correlate substantially with Self-confidence, Inquisitiveness and to a lesser extent with Analyticity. The second point is the consistently significant correlations between five CCTDI-UK sub-scales and the NEO Openness scale and the low correlations with the Systematicity sub-scale. The third noticeable point is the pattern and size of the inter-correlations between the CCTDI-UK sub-scales. There is an emergent pattern of two sub-groups: Included in the first group are the sub-scales Self-confidence, Inquisitiveness, Systematicity and in the second group is Analyticity, Open-mindedness, Truth-seeking. These patterns will be investigated at a later point in the thesis.

Table 5.18: Correlations between the sub-scales of the CCTDI-UK and the Openness scale of the NEO-PIR

Factor	Self-confidence N=372	Inquisitiveness N=372	Systematicity N=372	Analyticity N=372	Open-mindedness N=372	Truth-Seeking N=372	Open to Exper. N=95
Self-confidence	1	.534(**)	.397(**)	.182(**)	.109(*)	.124(*)	.450(**)
Inquisitiveness	.534(**)	1	.400(**)	.270(**)	.205(**)	.175(**)	.594(**)
Systematicity	.397(**)	.400(**)	1	.238(**)	.063	.087	.128
Analyticity	.182(**)	.270(**)	.238(**)	1	.328(**)	.402(**)	.341(**)
Open-mindedness	.109(*)	.205(**)	.063	.328(**)	1	.370(**)	.394(**)
Truth-Seeking	.124(*)	.175(**)	.087	.402(**)	.370(**)	1	.332(**)

** Correlation is significant at the 0.01 level (2-tailed).

* Correlation is significant at the 0.05 level (2-tailed).

5.7.1.4 Conclusions

The correlations with the NEOP-IR openness subscale provides strong evidence of convergent validity with this external criterion. The absence of correlations between the Systematicity sub-scale and the Truth-seeking, Open-mindedness and Openness scales may be expected as the rule based type of thinking associated with systematic thought is opposed to the flexible tolerance to different opinions associated with an open mind. Also the high number of inter sub-scale correlations provide further evidence of domain validity. Looking more closely at the correlations between critical thinking dispositions two groups of sub-scales show high inter-correlations suggesting a higher order component structure of critical thinking dispositions with one component composed of CT self-confidence, inquisitiveness and Systematicity and the other composed of Analyticity, Open-mindedness and Truth seeking. In short there are three arguments for the structure of critical thinking dispositions, univariate, bivariate and multivariate. The univariate solution suggests a general construct of critical thinking dispositions and is based on the general high degree of inter-correlation between all the sub-scales. A bivariate solution suggests two groups of sub-scales and is evidenced by the two highly inter-correlated groups described above. The multivariate solution suggests that each sub-scale should retain its own identity and this is based on the fact that although correlations are generally high they do vary in strength on a number of occasions. These theories are further examined empirically in the remaining studies of this thesis.

5.7.2 Study 4(b): Group Contrast Validity of the CCTDI-UK

5.7.2.1 Introduction

The following study investigates the group contrast validity of the CCTDI-UK with the test scores from a number of different populations. Group contrast validity is a further version of criterion validity, as a valid test should discriminate between groups of differing levels of development. Assuming higher education develops critical thinking dispositions then the scores on the CCTDI-UK should show differences between the students at different levels of their higher degree.

5.7.2.2 Method

Participants and Statistical analyses: There were three sets of participants used in Study 4(b). These three sets of participant scores were subjected to slightly different forms of statistical analyses. Data were collected on the CCTDI-UK in every cohort over the course of the research. In an effort to maximise the use of the data sets, several strategies were adopted. The main longitudinal sample had a high attrition rate and only 49 sets of scores remained for students who had participated at times 1, 2 and 3. However, more substantial data sets were available for times 1|2, 2|3 and 1|3 and these were also analysed.

The first data set consisted of 49 students from the longitudinal cohorts who had completed the CCTDI-UK at all three time-points (1st, 2nd and 3rd years of their degree). Their scores were analysed using a repeated measures ANOVA, which was used to investigate differences between sub-scale scores across each year of their degree. Significant differences were investigated with post-hoc tests. This analysis was entitled longitudinal (1).

The second set had three groups of participants, 70 students who completed the CCTDI-UK in both the 1st year and 2nd year of their degree, 70 who completed it in 2nd year and 3rd year and finally 62 who completed the inventory in 1st year and 3rd year. Their scores were analysed with paired sample t-tests because only two time-point variables apply to each of the three sets. This analysis was entitled split – longitudinal (1|2, 2|3, and 1|3).

The third set of data was completely cross-sectional and had a total of 370 students 119 entry (the same students from Cohort 2 in convergent study 5.7.2) and 63 Cohort 1 students (a total of 182 1st year students), 105 cohort 3 students (a total 105 year 2 students) and 58 exit students and 25 cohort 5 students (83 year 3 students). Tracked students that featured in this third set of participants only contributed one set of CCTDI-UK sub-scale scores. A one way between groups ANOVA was used to investigate the differences in the sub-scale scores between these cross-sectional cohorts of students. This analysis was entitled cross sectional.

These participant numbers are displayed by cohort in the following table.

Table 5.19: Summary of participants from each cohort

Study Analysis Title	No. from Cohort 1	No. from Cohort 2	No. from Cohort 3	No. from Cohort 4	No. from Cohort 5	Total No.
Longitudinal	49		49		49	49
Split-long 1,2	70		70			70
Split-long 2,3			70		70	70
Split-long 1,3	62				62	62
Cross-sectional	63	119	105	58	25	370

5.7.2.3 Results

Essentially there are five types of statistical analyses that might yield group contrast information for each of six sub-scales. The results are organised by sub-scale as each of the five analyses bear on the same question about group contrast validity with regard to each sub-scale. By examining the data-sets simultaneously the patterns across the different data sets can be easily detected.

CT Self-Confidence

Table 5.20 shows that there is a slight increase in the absolute mean between 1st year and 2nd year on self-confidence. Only one analyses from 5 showed a borderline (.049) significant difference between the years of study on the self-confidence scores. Although the ANOVA shows a significant increase in self-confidence the post-hoc tests show no significant difference between the individual levels.

Table 5.20: Summary data for group contrast validity for the CT Self-confidence sub-scale

Analyses	1 st year mean	2 nd year mean	3 rd year mean	Sig. difference
Longitudinal	3.94	4.03	3.93	.309
Split-long 1,2	3.92	3.97		.377
Split-long 2,3		4.06	3.98	.168
Split-long 1,3	3.98		3.96	.762
Cross-sectional	3.89	4.06	4.05	.049*

Inquisitiveness

Table 5.21 shows that there is an increase in the absolute mean in 2nd year on Inquisitiveness. Three analyses from five showed a significant difference between the years of study on the inquisitiveness scores. These differences occur mainly between 1st year and 2nd year. Only the cross-sectional analysis shows a significant difference. *Bonferroni Post hoc tests show that there is a significant difference between Level 1 and level 3 for inquisitiveness ($p < .001$).

Table 5.21: Summary data of group contrast validity for the Inquisitiveness sub-scale

Analyses	1st year mean	2nd year mean	3rd year mean	Sig. difference
Longitudinal	3.92	4.24	4.11	.004
Split-long 1,2	3.92	4.18		.001
Split-long 2,3		4.21	4.12	.168
Split-long 1,3	3.94		4.11	.059
Cross-sectional	4.01	4.19	4.40	.000*

Systematicity

Table 5.22 shows no significant differences between any of the years of study. There is a slight change in pattern between the longitudinal results and the cross sectional results as the absolute levels drop in the former and rise in the latter.

Table 5.22: Summary data of group contrast validity for the Systematicity sub-scale

Analyses	1st year mean	2nd year mean	3rd year mean	Sig. difference
Longitudinal	4.37	4.29	4.23	.338
Split-long 1,2	4.30	4.27		.743
Split-long 2,3		4.27	4.17	.139
Split-long 1,3	4.38		4.24	.114
Cross-sectional	4.22	4.30	4.38	.204

Truth-seeking

Table 5.23 shows a number of significant differences between the years of study. All analyses show an increase in the Truth-seeking scores over the three years in a step-like fashion. The major increase is shown to be between 2nd year and 3rd year.

*Bonferroni Post hoc tests show that there is a significant difference between Level 1 and level 3 for truth-seeking ($p < .001$). There is also a significant difference between level 2 and level 3 scores on truth-seeking ($p = .021$).

Table 5.23: Summary data of group contrast validity for the Truth-Seeking sub-scale

Analyses	1 st year mean	2 nd year mean	3 rd year mean	Sig. difference
Longitudinal	3.83	3.90	4.07	.035
Split-long 1,2	3.83	3.92		.249
Split-long 2,3		3.85	4.04	.024
Split-long 1,3	3.85		4.04	.010
Cross-sectional	3.87	3.91	4.17	*.002

Open-mindedness

Table 5.24 shows an overall pattern of decrease between 1st year and 2nd year then an increase back to original levels in 3rd year. The only significant difference significant difference occurs between 2nd year and 3rd year on the split-long 2|3 analysis.

Table 5.24: Summary data of group contrast validity for the Open-mindedness sub-scale

Analyses	1 st year mean	2 nd year mean	3 rd year mean	Sig. difference
Longitudinal	4.76	4.54	4.87	.066
Split-long 1,2	4.75	4.59		.193
Split-long 2,3		4.43	4.83	.006
Split-long 1,3	4.76		4.81	.566
Cross-sectional	4.76	4.55	4.81	.057

Analyticity

Table 5.25 shows generally a slight increase in absolute levels of analyticity over the three years. However these increases are not statistically significant in any of the five tests.

Table 5.25: Summary data of group contrast validity for the Analyticity sub-scale

Analyses	1st year mean	2nd year mean	3rd year mean	Sig. difference
Longitudinal	3.65	3.62	3.79	.187
Split-long 1,2	3.66	3.64		.094
Split-long 2,3		3.56	3.67	.138
Split-long 1,3	3.65		3.80	.094
Cross-sectional	3.65	3.61	3.79	.220

5.7.2.4 Conclusions

Looking at the absolute levels of the scores on the six dispositions it can be stated that there are no ceiling effects in their measurement. The maximum score achievable is six on the seven point scale used, and the mean scale scores fell in a range between 4.7 (Open-mindedness) and 3.7 (Analyticity). The order of sub-scales in relation to magnitude of scores from highest to lowest was. Open-mindedness (4.7), Systematicity (4.3), Inquisitiveness (4.1), Truth-seeking (4.0), Self-confidence (4.0) and Analyticity (3.7).

These results also show different discriminating properties of the various scales, which provide evidence for the developmental progress of students in these dispositions. Inquisitiveness and Truth-seeking generally show linear increases over the three years. Open-mindedness is a borderline case showing a decrease in second year and then an increase again in third year. Analysis, Self-confidence and

Systematicity showed no consistent significant changes. The observed changes may be due to one of two reasons. Either the scales have differing levels of group contrast validity or the actual levels of these dispositions are changing by different levels in the students throughout the course of their undergraduate degree. For the purposes of this chapter the results will be used principally as a measure of group contrast validity to add to the other measures of validity investigated in this chapter. Questions about the changing levels of critical thinking dispositions will be investigated in Chapter 7.

5.8 Study 5: Investigation of Higher Order Factors of Critical Thinking

Dispositions

5.8.1 Introduction

From the correlations in Table 5.18 it was felt that two higher order factors of critical thinking dispositions were emerging. The current study attempted to explore the factor structure to see if two higher order factors were present.

5.8.2 Method

Participants: There were 370 participants in this study. They were the same participants in the cross-sectional sample of the group contrast validity study (see 5.7.2.2)

Statistical Analysis: The factor analysis in this study is carried out on sub-scale means for each of the participants. Therefore there were only 6 input variables, i.e. for each participant. To test this, a maximum likelihood factor analysis was carried out. The rotation method was Direct Oblimin as the correlations in Table 5.18 showed a high degree of correlation between these sub-scales.

5.8.3 Results

The scree plot in Figure 5.2 and the cumulative variance explained (62%) clearly show a two factor solution for critical thinking dispositions as theorised from the correlations in Table 5.18. Table 5.26 shows the higher order factor structure for critical thinking dispositions. This structure matches that which emerges from the correlations in Table 5.18. The grouping of Inquisitiveness, Self-Confidence and Systematicity all load above .4 on factor 1 and henceforth is referred to as inquisitiveness because of the highest loading factor. The grouping of truth-seeking, open-mindedness and analyticity all load above .4 and are referred to as truth-

seeking because of the highest loading factor. The factor correlation matrix (Table 5.27) shows that there is a moderate correlation between the two factors.

Figure 5.2: Shows a Scree plot for the exploratory factor analysis on the subscale scores of the sample of participants described in 5.7.1

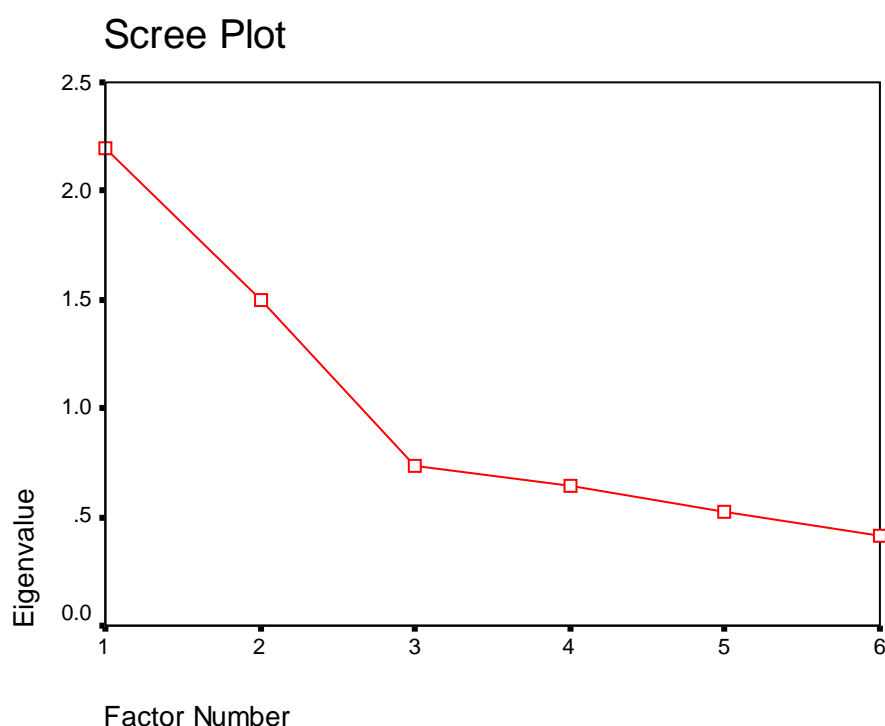


Table 5.26: Maximum Likelihood factor analysis structure matrix. Rotation method: Oblimin with Kaiser Normalization

	Factor	
	1	2
Inquisitiveness	.758	.160
Self-Confidence	.721	.261
Systematicity	.558	.086
Truth-seeking	.133	.734
Analyticity	.076	.577
Open-Mindedness	.256	.535

Table 5.27: Factor correlation matrix between the two emergent factors

Factor	1	2
1	1.000	.233
2	.233	1.000

5.8.4 Conclusions

The results of this study suggest that there are two higher order factors of critical thinking dispositions. The emergent factors are composed of the two highly correlated groups identified in Study 4(a). Therefore the bi-variate solution proposed in Study 4 and supported in the current study warrants further investigation. Chapter 7 and 8 make further comments on the bi-variate solution of critical thinking dispositions.

5.9 Discussion

The main theme of the chapter was to assess the reliability and validity of scores on the CCTDI. The confirmatory factor analysis in Study 1 found that the construct validity of the CCTDI-UK was suspect and concluded that the next step was to use an exploratory factor analysis to investigate its structure. This was carried out in Study 2 and a reliability analysis was also used. The factor analysis showed a seven factor structure within the test. However, the factor analysis found a number of the original items were problematic, i.e., they were low loading, cross loading or poorly worded. As a result they were removed from the scale by the experimenters. The remaining items were retained but often within different sub-scales than they were originally positioned. Further items, in fact a complete sub-scale (Maturity), were

removed through the reliability analysis carried out in the study. The revised test produced by restructuring and item removal was named the CCTDI-UK because it was refined version of the CCTDI prepared for use within UK populations. The confirmatory factor analysis carried out on the revised scale in Study 3 showed an improvement in construct validity. The reliability analysis comparison between the CCTDI and the CCTDI-UK, also included in this study, also showed significant improvement in the psychometric strength of the revised scale. Study 4(a) showed correlations between the CCTDI-UK sub-scales and the Openness scale from the NEO PI-R. The results suggested that there are three possible solutions to the structure of critical thinking dispositions measurement, i.e., uni-variate, bi-variate and multivariate. These measurement models are discussed further in Chapters 7 and 8. Study 4(b) showed that the Truth-Seeking and Inquisitiveness sub-scales showed the most significant increases over the test period, i.e., three years of a psychology degree. Open-mindedness showed a decrease in second year and a recovery in the third year of university. The other scales showed no significant changes. The factor analysis in Study 5 supported the bi-variate solution of critical thinking dispositions.

The remainder of this discussion attempts to aggregate the evidence for reliability and validity for each of the subscales thus providing an overall picture of their psychometric strength as individual scales. This is done with a rank based system that places the subscales in order 1-6 for the four indicators of psychometric strength, i.e. internal consistency, construct validity, convergent validity and group contrast validity. The following Table 5.28 summarises these ranks and the overall psychometric strength scores are visually displayed in Figure 5.3. The sub-scales are ranked by the following criteria:

Reliability – Highest Cronbach's Alpha

Construct Validity – Highest significance level of goodness of fit model

Convergent Validity – Highest correlation co-efficient mean with other sub-scales

and the NEO PI-R

Group contrast Validity – Most significant discriminations

Overall rank score – Summary score compiled from ranks on the various reliability

and validity tests using the following point system; 6 points for ranked first, 5 for

ranked 2 etc.

Table 5.28: Rank based summary of the psychometric strength of the CCTDI-UK sub-scales

Indicator	Inquisitiveness	Truth-seeking	Self Confidence	Open-mindedness	Analyticity	Systematicity
Reliability	2	5	1	6	3	4
Construct Validity	6	3	4	1	5	2
Convergent Validity	1	3	6	4	2	5
Group contrast Validity	2	1	3	3	5	5
Overall Rank Score	17	16	14	14	13	12

Figure 5.3: Graphical representation of the overall psychometric strength rank score of the CCTDI-UK sub-scales

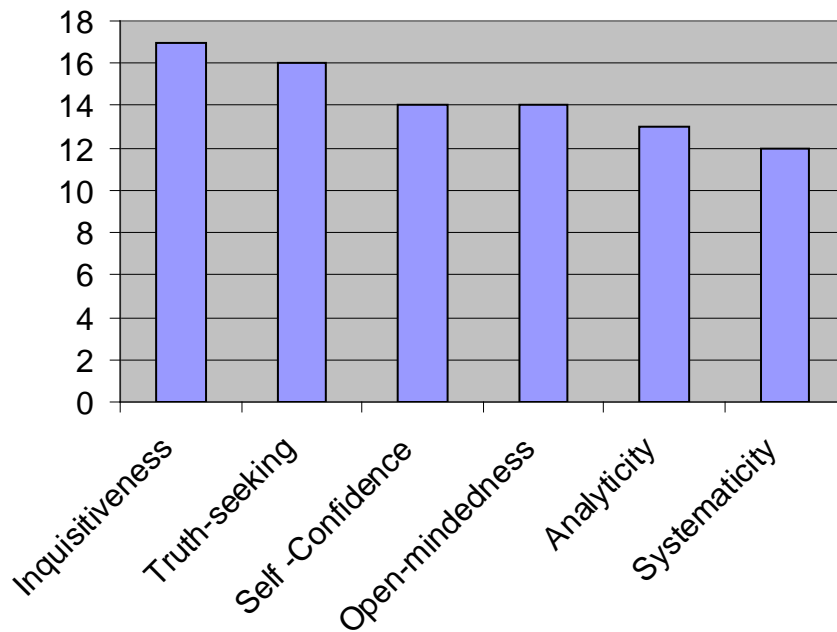


Table 5.28 and Figure 5.3 are an attempt to aggregate the sources of reliability and validity and to estimate the psychometric strength of the CCTDI-UK sub-scales. Looking at Figure 5.3 it can be seen that the order of psychometric strength for these sub-scales is (strongest to weakest psychometrics) Inquisitiveness, Truth-seeking, CT Self-confidence, Open-mindedness, Analyticity and Systematicity. However, these numbers should be treated with caution. From the distribution of ranks in Table 5.28 it can be seen that the variation of ranks is quite large and the range of overall scores is relatively narrow, and all the scales have their particular strengths and weaknesses and vary within a narrow range. In addition, the tests of reliability and validity have been given equal weightings in the aggregate and they probably do not have equal value. For example, reliability is a prerequisite for all forms of validity yet reliability has been offered an equal standing in the summary. Therefore, the numbers should not be interpreted at face value. Nevertheless the exercise does indicate that generally good level of psychometric strength for all these sub-scales and that Truth-Seeking and Inquisitiveness are particularly strong measures.

Chapter 6

A Psychometric Analysis of Critical Thinking Skills: The California Critical Thinking Skills Test (CCTST)

6.1 Introduction

The main focus of this chapter is the reliability and validity of critical thinking skills measurement. The instrument central to this investigation is the California Critical Thinking Skills Test (CCTST). This scale has 34 items which are divided into three or two sub-scales depending on which of the two available characterizations, offered in the test manual, is applied. The first characterization features three sub-scales, called Analysis, Inference and Evaluation. These sub-scales are derived from the theory in the Delphi Report (Facione, 1991) which is outlined in Chapter 3. The other characterization has two sub-scales, Inductive and Deductive reasoning. The justification for offering these sub-scales in the CCTST manual is that they are more traditional categories. Only 30 items were placed on these two traditional scales. The following paragraphs provide a short description of each scale and a typical item is also provided.

Analysis: Facione et al. (1998, p.5) state that the analysis construct has a dual meaning. Firstly, they suggest it means ‘to comprehend and express the meaning or significance of a wide variety of experiences, situations, data, events, judgements, conventions, beliefs, rules, procedures or criteria.’ They then suggest it also means ‘to identify the intended and actual inferential relationships among statements, questions, concepts, descriptions or other forms of representation intended to express beliefs, judgements, experiences, reasons, information or opinions.’

Analysis item

“Not all the candidates are qualified to serve,” expresses the same idea as:

A= None of the candidates are qualified to serve.

B= Some candidate is not qualified to serve.

C= Someone qualified to serve is not a candidate.

D= All candidates are not qualified to serve.

Answer = B

Evaluation: Again, Facione et al. (1998, p.5) state that Evaluation has a double meaning. They state it means ‘to assess the credibility of statements or other representations which are accounts or descriptions of a person’s perception, experience, situation, judgement, belief or opinion; and to assess the logical strength of the actual or intended inferential relationships among statements, descriptions, questions or other forms of representations. They suggest it also means ‘to state the results of one’s reasoning; to justify that reasoning in terms of the evidential, conceptual, methodological, criteriological and contextual considerations upon which one’s results are based; and to present one’s reasoning in the form of cogent arguments.’

Evaluation Item

Passage: “Terry don’t worry about it. You’ll graduate someday. You’re a college student Right? And all college students graduate sooner or later.” Assuming all the support statements are true, the conclusion

A= could not be false.

B= is probably true, but may be false.

C= is probably false, but may be true.

D= could not be true.

Answer = A

Inference: Facione et al. (1998, p.5) state that Inference means ‘ to identify and secure elements needed to draw reasonable conclusions; to form conjectures and hypotheses, to consider relevant information and to deduce the consequences flowing from data, statements, principles, evidence, judgements, beliefs, opinions, concepts, descriptions, questions or other forms of representation.’

Inference Item

Consider this group of statements: “Nero was emperor of Rome in the first century AD. Every Roman emperor drank wine and did so using exclusively pewter pitchers and goblets. Whoever uses pewter, even once, has lead poisoning. Lead poisoning always manifests itself through insanity.’

Which of the following must be true if all of the above are true?

A= Those who suffer from insanity used pewter at least once.

B= Whatever else, emperor Nero was certainly insane.

C= exclusive use of pewter was a privilege reserved for Roman emperors.

D= Lead poisoning was common among the citizens of the Roman Empire.

Answer = B

Inductive Reasoning: Facione et al. (1998, p.6) states ‘Inductive reasoning as used in the CCTST sub-scale means an argument’s conclusion is purportedly warranted, but not necessitated, by the assumed truth of its premises. Scientific confirmation and experimental disconfirmation are examples of inductive reasoning.

Inductive reasoning item

For this question use this fictitious case: ‘In a study of high school students at Mumford High, it was found that 75% of those students who drank two or more beers each day for a period of 60 days experienced measurable liver function

deterioration. That these results could have occurred by chance was ruled out experimentally with high levels of confidence.'

If true, the Mumford High information would confirm that

A= Drinking is statistically correlated with liver deterioration in adolescents.

B= Drinking causes liver deterioration in adolescents.

C= Sex is not a factor in the relationship between alcohol and liver deterioration.

D= The researcher had a personal reason to want to prove young people that they should not drink.

E= The drinking age laws are out of date and should be changed.

Answer = A

Deductive Reasoning: Facione et al. (1998, p.6) state 'Deductive reasoning as used in the CCTST sub-scale means the assumed truth of the premises purportedly necessitates the truth of conclusion. Not only do traditional syllogisms fall within this category, but algebraic, geometric and set-theoretical proofs in mathematics (including 'mathematical induction') also represent paradigm examples of deductive reasoning.

The example Analysis and Evaluation items, outlined above, also load on the Deduction sub-scale.

Two other tests that feature in this chapter are the short form of Raven's Advanced Progressive Matrices (RAPM-sf) and A-level results. The RAPM-sf is a test on non-verbal intelligence that correlates more highly with g than any other form of intelligence test (Cooper 1998). This is multiple choice test that provides eight

options in the form of shapes that must be placed in the space created in a number of shape sequences in the above diagram. The participant must look up and down the sequences to choose the most appropriate shape to continue that sequence pattern. (see Appendix 6.1 for an example). There are twelve items in this test which increase in difficulty throughout the test. The test has a ten minute time limit. A-levels are the standard tests of academic knowledge that are most often used for selecting students into UK universities.

Little independent research exists into the reliability and validity of the critical thinking skill measurement. Most of what is known can be found in the manuals of the tests that purport to measure critical thinking and a few associated articles by their publishers. This evidence is provided in matrix 2 and 3 (Appendix 3.1 and 3.2) which show the reliability and validity estimates from research carried out with this test. This information is summarised at the end in Chapter 3 in Figure 3.1 and 3.2. Looking at criterion validity and reliability information for the CCTST reported Appendix 3.1, it could be suggested that this information is neither substantial nor wide in its scope of populations. The purpose of this chapter is to provide additional evidence for reliability and validity of the CCTST, particularly for measuring critical thinking skills in UK populations.

The following chapter reports two studies which investigate the psychometrics of critical thinking skills testing in undergraduate psychology students. The first study (6.2) reports reliability information for the CCTST. Kuder Richardson-20 scores are used as the estimates of reliability of the CCTST and its sub-scales. A revised version of the CCTST called the CCTST-UK is proposed from the results in this

study. The second study (6.3) reports on the validity of the newly proposed CCTST-UK. The type of validity evidence provided in this study is criterion validity. Firstly, this study details empirical evidence of the convergent validity of the CCTST-UK. The correlations presented show the relationships between the sub-scales of the CCTST-UK, a measure of general intelligence (Raven's Advanced Progressive Matrices Short form – RAPMsf), and a measure of academic achievement (A-levels). Next, the study offers evidence of criterion validity of the CCTST using contrasted groups. This evidence shows differences in critical thinking skills between undergraduates entering university and students who had almost finished their psychology degree. The chapter ends with a brief discussion of the experimental findings.

6.2 Study 6: The internal consistency of the CCTST

6.2.1 Introduction

The purpose of the study is to provide evidence for the internal consistency of the CCTST on a sample of undergraduate psychology students in a UK higher education institution. The main aim is to check the consistency of the scale and to adapt the measure to improve it, if necessary.

6.2.2 Method

Participants: There were 228 participants in this study who had completed the CCTST. They were composed of 184 students in the first year of their degree (cohort 2) and 44 3rd year students (cohort 4). There were 188 female and 40 male participants. The 1st year data was collected in a specially designed laboratory class for 1st year students. The class was run with the same content, at the same time of day, on each of four weekdays. The instructions before the test were given as outlined in the CCTST manual (Facione et al., 1998). The students were asked to put their names on the test paper as the experimenters wished to match their scores with scores on other tests in the study. Confidentiality was assured.

Statistical Analysis: A Kuder Richardson-20 item analysis was used to initially check the internal consistency of the test. Kuder Richardson-20 is used for checking the reliability of dichotomous items. The reliability analysis was also used to eliminate unreliable items from the various sub-scales in this test.

6.2.3 Results

Table 6.1 shows Kuder Richardson-20 scores for the sub-scales (Evaluation, Inference, Analysis, Deductive Reasoning and Inductive Reasoning) and total score of the CCTST. The table also shows the number of items in each of these divisions. The data suggests that the Analysis scale should be removed from the Delphi sub-scale characterisation because of its negligible Kuder Richardson-20 score. The other scales have a relatively low Kuder Richardson-20 score and thus suggest suitability for an item analysis to improve the internal consistency.

Table 6.1: Kuder Richardson-20 scores and item numbers for the CCTST total and sub-scales (Evaluation, Inference, Analysis, deductive reasoning and inductive reasoning)

Category	Items	Number of Items	Kuder Richardson-20 (N=228)
CCTST	1-34	34	.57
Sub-Scale Category 1			
Evaluation	1,2,3,4,25,26,27,28,29,30,31,32,33,34	14	.51
Inference	14,15,16,17,18,19,20,21,22,23,24	11	.40
Analysis	5,6,7,8,9,10,11,12,13	9	.00
Sub-Scale Category 2			
Deductive Reasoning	1,2,4,5,6,8,9,14,15,16,17,18,19,22,23,27	16	.47
Inductive Reasoning	3,13,20,21,24,25,26,28,29,30,31,32,33,34	14	.41

Table 6.2 below shows the new Kuder Richardson-20 scores of the retained sub-scales of the CCTST (analysis removed because of negligible reliability) when the unreliable items had been removed. This table also shows the items that were removed and the number of items that remained. Generally, there has been a reduction in items and an improvement in reliability for all sub-scales and the two category totals (Delphi characterisation and Traditional characterisation) of the CCTST.

Table 6.2: Kuder Richardson-20 scores, items removed and number of items for the two new CCTST versions and revised sub-scales (Evaluation Inference, Deductive and Inductive reasoning)

Sub-scale	Items Removed	Number of Items		Kuder Richardson-20 (N=228)
		Before	After	
Sub-Scale Category 1				
Evaluation-UK	3,28,30,33	14	10	.53
Inference-UK	16,17,20	11	8	.46
Combined total (CCTST-UKa)	3,5,6,7,8,9,10,11,12,13,16,17,20,28,30,33	34	18	.63
Sub-Scale Category 2				
Deductive Reasoning	5,8,9,16,17	16	11	.54
Inductive Reasoning	3,13,21,24,33,30	14	8	.46
Combined Total (CCTST-UKb)	3,5,8,9,13,16,17,21,24,30,33	30	23	.60

6.2.4 Conclusions

The Kuder Richardson-20 scores show marginal reliability for the sub-scales and scale totals. They are slightly lower than the Kuder Richardson-20 scores produced by the CCTDI-UK (see Study 3). However, Nunally (1978) suggests that Kuder Richardson-20 scores are generally lower than Cronbach's Alpha scores because of the reduced range in the correlation co-efficient. Table 6.2 also suggests that subscale category 1 (Delphi characterisation) is the best format for measuring critical thinking skills because it has similar internal consistency but better content validity based on the Delphi report. Therefore version (a) of the CCTST-UK has been chosen for further analysis in this thesis. This revised version of the CCTST was re-named as the CCTST-UK for ease of reporting.

6.3 Study 7: Criterion validity of the California Critical Thinking Skills Test-UK

6.3.1 Introduction

The criterion validity information in this study on the CCTST-UK is produced by two methods namely, correlations with other tests and group contrasts. The correlations are between the CCTST total score, CCTST Evaluation sub-scale score, CCTST Inference sub-scale score, a measure of non-verbal intelligence (RAPM-sf) and academic achievement in the form of A-levels. Theoretically all these measures cover an aspect of cognitive ability and/or attainment therefore there should be a number of significant correlations displaying convergent validity.

The group contrasts are between the entry level (1st year students) and exit level (3rd year students). Assuming that higher education develops critical thinking skills, a valid critical thinking skills measure should show a significant improvement between degree entry and exit students.

6.3.2 Method

Participants: There were 228 participants in this study that had completed the CCTST. They were made up of 184 entry students (cohort 2) and 44 exit students (cohort 4).

Not all of these students feature in all of these analyses as different numbers have completed the various measures (see cohort descriptions in 4.3.2).

Statistical analyses

The convergent validity was investigated using Pearson's correlations (table 6.3) and the groups were contrasted using independent t-tests.

6.3.3 Results

Table 6.3 shows the mean scores on each of the various measures. There is a max score of 18 for the CCTST-UK, 10 for the evaluation sub-scale, 8 for the inference sub-scale, 30 for A-level points and 12 for the RAPM-sf. These measures all show a normal distributions except RAPMsf which is skewed to the top end. A ceiling effect occurs here as university students are generally score highly on the short form of the Raven's Advanced Progressive Matrices.

Table 6.3: Descriptive statistics for each of the measures used in Study 2

	N	Mean	Std. Deviation
CCTST-UK Evaluation (10 items)	228	4.61	2.06
CCTST-UK Inference (8 items)	228	3.56	1.65
CCTST-UK Total Score (18 items)	228	8.16	3.08
A-level Points	168	23.95	3.79
RAPMsf	224	10.61	1.3

Table 6.4 below provides several points of note

- The sub-scales are highly correlated with total score which is not surprising, as these two scores make up the total score.
- The evaluation and inference skills are significantly correlated ($r=.364$). This suggests that the two scales may be added together to provide a total score for critical thinking skills (this issue is discussed later in the thesis p.184). Furthermore, ability tests traditionally show additive potential unlike personality or dispositional tests.

- Also the CCTST-UK total and sub-scales are correlated with the non-verbal intelligence measure and slightly less strongly with A-levels. The inference sub-scale follows the same pattern but is less evident than the evaluation sub-scale.
- A-levels are not correlated with the non-verbal intelligence measure ($n = 173$, $r = -.05$, $p = .849$).

Table 6.4: Pearson's Correlations between Sub-scales of the CCTST-UK, Total Score CCTST, A-level Points and RAPMs Score

		CCTSTUK Total N=228	CCTST-UK Evaluation Sub-scale N=228	CCTST-UK Inference Sub-scale N=228	Total A-Level points N=168	Raven's Score N=224
CCTSTUK Total	r	1	.866(**)	.781(**)	.230(**)	.284(**)
CCTSTUK Evaluation Sub-scale	r	.866(**)	1	.364(**)	.177(*)	.223(**)
CCTSTUK Inference Sub-scale	r	.781(**)	.364(**)	1	.203(**)	.250(**)

** Correlation is significant at the 0.01 level (2-tailed).

* Correlation is significant at the 0.05 level (2-tailed).

Table 6.5 below provides evidence for criterion validity of the CCTST-UK and its sub-scales. Group contrast as seen in the last chapter (Study 4(b)) is a method of identifying criterion validity. The two groups used in this study are students in their 1st year at university (entry sample) and students in their 3rd year of higher education study (exit sample). Assuming critical thinking skills are developed in higher education then these tests should identify differences between these two groups.

The means show increases in the CCTST-UK and its subscales between entry and exit students. These increases are significant except on the inference sub-scale

(which is approaching significance.) The RAPM-sf scores show no significant difference between entry and exit students. Furthermore there are no differences between the A-level score of the exit and entry sample. Therefore it can be stated that the groups are matched for non-verbal intelligence and academic achievement.

Table 6.5: Means and Independent t-tests Between Entry and Exit Students on the CCTST-UK, CCTST-UK Subscales and RAPMsf

	Cohort	N	Mean	Std. Deviation	Std. Error Mean	t	df	Sig.
RAPMSF TOTAL	Entry	185	10.59	1.30	.10	-.22	233	.830
	Exit	50	10.64	1.41	.20			
CCTST-UK Evaluation (10 items)	entry	184	4.39	2.02	.15	-3.27	226	.001
	Exit	44	5.50	2.01	.30			
CCTST-UK Inference (8 items)	Entry	184	3.45	1.65	.12	-1.82	226	.070
	Exit	44	3.95	1.60	.24			
CCTST-UK Total Score (18 items)	Entry	184	7.85	3.07	.23	-3.16	226	.002
	Exit	44	9.45	2.80	.42			
A-levels	Entry	132	23.98	3.93	.34	.234	64.85	.815
	Exit	36	23.83	3.90	.55			

6.4 Discussion

Overall the internal consistency of the CCTST was poor. This prompted the production of a revised scale called the CCTST-UK. Although reliability of the CCTST-UK was not raised to an ideal level by item analysis, it was increased and a substantial number of unreliable items were removed including a complete scale of the CCTST.

The correlations with other tests of cognition and attainment provided evidence of good convergent validity. There is a slightly higher correlation between CCTST-UK scores and RAPMs than A-levels. The reason for this is that the CCTST is not embedded in discipline knowledge similar to the RAPMs but unlike A-levels where assessment is almost completely based on discipline knowledge. Although the CCTST-UK and its sub-scales show moderate correlation with non-verbal intelligence and academic attainment in the form of A-levels it could be suggested that they are identifying particular qualities of the students that may need further investigation. This claim is continued in the next chapter when critical thinking skills are explored in the context of higher education attainment.

Looking at the absolute change in the scores on the test it can be stated that there are no ceiling effects occurring in relation to the CCTST-UK total score and both sub-scale scores. In fact mean participant scores are generally mid range. For example there are 10 items in the Evaluation sub-scale, thus a maximum score of 10 achievable and the entry students have a mean score of 4.39 and the exit students have a mean of 5.5 (see Table 6.5). Overall the scales also show good criterion validity using the method of group contrast. If students are developing critical

thinking skills during the course of their higher education then the CCTST-UK is sensitive to this development.

Chapter 7

Using Scores on the CCTST-UK and CCTDI-UK to Predict Degree Attainment

7.1 Introduction

There are two main purposes of this chapter. The first is to provide some further criterion validity for the CCTDI-UK and CCTST-UK in the form of predictive validity. Concurrently, the chapter investigates an applied issue about the selection of students for higher education. Scores on the two revised tests produced in Chapters 5 and 6 are assessed for their ability to predict a student's level of attainment in a higher education programme. The major assumption in this investigation is that higher education assessment is sensitive to some aspects of critical thinking albeit assessed in the context of disciplinary knowledge. Therefore if these critical thinking tests are reliable and valid they should correlate with and predict attainment in higher education. Furthermore, if these predictive qualities were impressive, especially at the level of entry students, then they could be potential methods for use in selecting students for higher education.

The prediction of student's attainment is a useful goal because it provides universities with a potential indicator by which they can select students who will thrive in their degree programmes. The first section (7.1) in this chapter provides some detail about issues surrounding selection for higher education. The chapter then reports the predictive validity of scores on the refined versions of the CCTST and the CCTDI for predicting degree attainment in higher education (7.2, Study 8). The particular focus of the study is to identify the predictive strength of critical thinking tests when completed on entry to higher education, thus providing evidence for the

usefulness of these tests as higher education selection instruments. The last part of this chapter (7.3, Study 9) returns to questions of predictive validity and examines critical thinking constructs, at final year of study (exit rather than entry) that predict degree attainment. Attempts are then made to build correlative pathways between entry and exit critical thinking constructs as well as indicators of degree attainment.

7.2 Higher Education Admissions: A Role for Critical Thinking Measurement?

Assessment and selection for Higher Education is a hotly debated topic. As there are limited places in higher education, institutions must operate some form of selection criteria. Much public, political and media attention has focused on the fairness and validity of the current practices used to select students for third-level education. Furthermore there is growing concern among many individuals and institutions that A-levels are not sufficiently discriminative to select students for higher education places.

As a result the UK Government established an advisory group to look into the debate. The Admissions to Higher Education Steering Group (AHESG), was headed by Professor Steven Schwartz, Vice-Chancellor of Brunel University. This steering group investigated a number of issues including; the current admissions practices, fairness of those practises and possible future options for admissions practises. One of the options suggested was aptitude testing, of which critical thinking tests are an example.

Selection for Higher Education varies greatly among global national education systems. There are number of methods that are currently being used including A-

levels, GCSEs, AS-levels, Advanced Extension Awards (AEAs), vocational and access qualifications, baccalaureate style qualifications (Irish Leaving Certificate, Scottish Higher Exams, Welsh Baccalaureate), aptitude testing, interviews, portfolios and auditions, school performance, contextual factors, special credit, accreditation of prior experiential learning (APEL), specialist subjects, clearing and random selection. This chapter focuses on two of these methods namely A-levels which is the method mostly used by third level institutions. Within the context of this thesis the focus will be on A-levels and components of critical thinking skills and dispositions.

There is increasing interest in aptitude testing as a possible selection device. Aptitude tests have been used for selection to competitive higher education courses like medicine (Kreiter, Stansfield, James and Solow, 2003; Searle and McHarg, 2003; Higgins and Sun, 2002), and are being piloted for more general use by universities to select students. For example, 1551 students applying to Cambridge in 2003/4 took a thinking skills tests developed the Oxford Cambridge Local Examination Syndicate (UCLES, 2004).

Another important question is how to gauge success in a degree programme. In this context the word 'Success' is used as an umbrella term to include a number of outcomes namely, degree completion, good degree result and post degree societal contribution (AHESG, 2003). Responses to the consultation paper by the AHESG suggest that a combination of these factors should be used to gauge 'Success'.

However, it is much easier to gauge 'success' by examining degree or post-graduate degree results than to assess the more general impact that a graduate might have on

society. Within the limits of this thesis, we will confirm the questions of predictive validity, of both A-levels and critical thinking measures, using the success indicator of final undergraduate degree results (and sub-components). Critical thinking was chosen for investigation as a potential indicator of 'success' because of the benchmarking statements outlining critical thinking as an important goal in higher education (see section 3.2.2). Furthermore, many educators have emphasised that students who are good critical thinkers will achieve success in higher education (Barbanel, 1987; Chaffee, 1997; Elder and Paul, 2003; Ennis, 1996a; Feldt, 1989; Higbee, 2003; Higbee and Dwinell, 1998; James, 2002; Meyers, 1987; Paul and Elder, 1996, 2002, 2003).

7.3 Study 8: Using 1st year Levels of Critical Thinking to Predict Degree Attainment

7.3.1 Introduction

This study initially looks at the correlations between participant entry level scores on the CCTST-UK, CCTDI-UK, RAPM-sf and A-levels, with a number of degree attainment indicators, i.e. their overall degree average and their average marks in year 1, 2 and 3 of their degree. These correlations were carried out to identify input variables for the subsequent regression analysis to find predictors of degree attainment, as any measures correlating with the degree attainment indicators have potential as predictors. These potentially predictive test scores were then entered into regression analysis to investigate their predictive strength when A-levels were controlled for in the regression model. A-levels were controlled because the rationale was to investigate how much additional variance these measures can predict additionally to A-levels when A-levels scores are known, thus, the analysis attempts to measure a real world scenario where these aptitude tests might be used alongside A-levels to select students for higher education.

7.3.2 Method

Participants: The entry (N=190) sample of students (cohort 2) participated in this study as they have scores for A-levels, CCTST-UK, CCTDI-UK, RAPMsf and degree marks. However not all these students have scores for each of these tests and numbers vary for the different statistical analyses in this chapter.

Statistical Analyses: The four degree attainment indicators (degree averages, 1st, 2nd, 3rd year averages) were correlated with the subscales of the two critical thinking tests, A-level points and RAPM-sf. These correlations were used to identify potential

predictors and any measures that significantly correlated were entered into a hierarchical regression which controlled for A-level scores.

7.3.3 Results

Table 7.1 shows correlations between the various measures of degree attainment and the scores on the CCTST-UK total, the two CCTST sub-scales (Evaluation and Inference), A-level points and RAPM-sf. The purpose of these correlations is to select tests for the regression analysis (Table 7.2). There were no significant correlations with scores on the CCTDI-UK sub-scales and the degree attainment indicators; so they were not included in this table and therefore will not be considered for regression analysis at this stage. There are a number of points to be considered from the table:

Table 7.1: Correlations between Degree marks (L1, L2 and L3), A-Levels, RAPMsf, CCTST scores (evaluation subscale, inference sub-scale and total score)

		Degree Average	3 rd year average	2 nd year average	1 st year average	A-level points	1 st year RAPM- sf scores	1 st year CCTST- UK Evaluation score	1 st year CCTST- UK Inference score	1 st year CCTST- UK total score
Degree Average	Pearson Correlation	1	.937(**)	.623(**)	.735(**)	.323(**)	.081	.153	.250(**)	.241(**)
	N	116	106	116	116	94	116	116	116	116
3 rd year average	Pearson Correlation	.937(**)	1	.606(**)	.535(**)	.276(*)	.011	.113	.282(**)	.236(*)
	N	106	106	106	106	85	106	106	106	106
2 nd year average	Pearson Correlation	.623(**)	.606(**)	1	.622(**)	.365(**)	.088	.188(*)	.222(*)	.248(**)
	N	116	106	116	116	94	116	116	116	116
1 st year average	Pearson Correlation	.735(**)	.535(**)	.622(**)	1	.429(**)	.100	.187(*)	.175	.220(*)
	N	116	106	116	116	94	116	116	116	116

** Correlation is significant at the 0.01 level (2-tailed).

* Correlation is significant at the 0.05 level (2-tailed).

- As expected, there are high significant correlations between the various degree attainment indicators.
- Of the non-degree higher education linked assessments, A-level points show the highest correlations with the degree attainment.
- There are no significant correlations between entry level scores on the RAPM-sf test of non-verbal intelligence and any of the degree attainment indicators.
- There are low and marginally significant correlations between the 1st year students' scores on the Evaluation sub-scale of the CCTST-UK and both 1st year and 2nd year exam marks but not with third year exam marks and overall degree marks.
- There are moderate and highly significant correlations between the 1st year students scores on the Inference sub-scale of the CCTST-UK and 2nd year marks, 3rd year and overall degree marks but not with first year exam marks although it approaches significance.
- The total score of the CCTST-UK correlates with all the degree attainment indicators.

Table 7.2 below shows summary results of ten hierarchical regressions. The dependant variable in each case was one of the measures of degree attainment (1st year, 2nd year, 3rd year or overall degree marks). The independent variables consisted of the score of one of the four non-higher education related assessments completed by entry level (1st year) students namely; A-levels, CCTST-UK total, CCTST-UK inference, CCTST-UK evaluation. Only A-levels and the various skills test scores were entered into the regression because they were found to correlate with the degree attainment indicators in Table 7.1. Entry level scores on the RAPM-sf and CCTDI-UK sub-scales do not feature in any of the regressions because they showed no significant correlations with the degree attainment measures. Therefore, it can be said that these measures were not significant predictors of degree attainment.

The first column in Table 7.2 ‘Dependant outcome measure’ is the degree attainment indicator that is predicted by the ‘Predictors’ shown in the second column. The third column entitled ‘R square for A-level’ shows the percentage of the variance predicted by A-levels. The first block in each regression consisted of the independent variable ‘A-levels’, which meant that A-levels acted as a control. Therefore, the third column ‘R-square change’, indicates the predictive strength of the other independent variables, i.e., the predictive value that each I.V. had in addition to that of A-levels. The last column ‘significance of change’ shows whether the additional prediction is significant. In ecological terms these results show the validity for predicting degree outcome these tests would have if they were used alongside A-levels on entry level students.

Table 7.2: Summaries of hierarchical regressions showing r square change and sig. of change for each of the predictors when A-levels are controlled

Dependant /Outcome Measure	Predictor	R square for A-level * = sig. predictor	R square change (additional prediction)	Sig. of change
Degree Average	Entry (N=93) CCTST-UK Inference	.104*	.038	.048
	Entry (N=93) CCTST-UK Total	.104*	.037	.050
Level 3 Average	Entry (N=84) CCTST-UK Inference	.076*	.056	.024
	Entry(N=84) CCTST-UK Total	.076*	.034	.084
Level 2 Average	Entry (N=93) CCTST-UK Evaluation	.133*	.040	.038
	Entry (N=93) CCTST-UK Inference	.133*	.031	.070
	Entry (N=93) CCTST-UK Total	.133*	.052	.018
Level 1 Average	Entry (N=93) CCTST-UK Evaluation	.184*	.030	.065
	Entry (N=93) CCTST-UK Inference	.184*	.006	.433
	Entry (N=93) CCTST-UK Total	.184*	.023	.105

There a number of points worth noting from this table.

- A-levels account for approximately 10% of the variance when predicting degree attainment (highest = 18.4% of 1st year marks, lowest 7.6% of 3rd year marks).
- A-levels were significant predictors in all instances.
- Entry level total score on the CCTST-UK was a significant predictor of second year marks ($p=.018$) accounting for just over 5% of the variance and

marginally significant for overall degree average ($p=.05$) accounting for 3.7% of the variance.

- Entry level scores on the CCTST-UK Inference sub-scale were significant predictors in two instances. These scores significantly predicted 5.6% ($p=.024$) of the variance (the highest prediction after A-levels) for level 3 marks, compared to A-levels which predicted 7.6% of the variance here. The inference scores also significantly predicted 3.8% of the variance ($p=.048$) of overall degree mark.

Entry level scores on the CCTST-UK Evaluation sub-scale were significant predictors in one instance. These scores significantly predicted 4% ($p=.038$) of the variance for level 2 marks.

7.4 Study 9 Using 3rd Year Levels of Critical Thinking to Predict Degree Attainment

7.4.1 Introduction

The study outlined below identifies measures of critical thinking which can help to predict degree outcome and third year average exam marks when completed by students in the third year of their degree. The rationale behind this study is that if third year levels of certain critical thinking components predict degree outcome then a measurement model of 3rd year critical thinking can be produced (this is done in the discussion of this chapter, Figure 7.1 and 7.2). Furthermore, it would be beneficial to know what 1st year measures predict 3rd year levels of these degree attainment predictors where possible. Consequently, there are two parts to this study; as the first part identifies those critical thinking components measured in third year that predict degree attainment; the second part takes these identified components and, where the data set permits investigates which entry level scores predict their third year levels.

Part 2 of the study investigates the 1st year critical thinking dispositions validity for predicting 3rd year scores on critical thinking Inquisitiveness. The results suggest that more than any critical thinking disposition inquisitiveness in a third year student showed the highest predictive validity for degree attainment. Therefore it was felt that it would be useful to predict third year level of this disposition in its own right.

7.4.2 Method

Participants: There are two parts to this investigation and two sets of participants. The first sample is made up of 31 students in their 3rd year of degree who had completed both revised critical thinking tests (CCTST-UK and CCTDI-UK) and had

degree results available (taken from cohort 4). These 31 students completed the tests during a special class on critical thinking measurement within a 3rd year elective model on cognition and instruction. They were instructed as per the manuals of the CCTST and CCTDI (Facione et al., 1998 and 2000 respectively).

The second sample had 62 participants. These participants were taken from the longitudinal sample (cohort 1 and 3) and included any students who had completed the CCTDI-UK in their 1st year and again in their 3rd year. The tests were administered as instructed by the CCTDI manual (Facione et al., 2000).

Statistical analyses: The first part of this study correlated the scores of the CCTST-UK and the CCTDI-UK with degree result to identify input variables for regression analysis. A stepwise regression analysis was carried out with the dependant variable being degree result and the independent variables being scores on critical thinking tests that correlated significantly with degree result and third year marks. A stepwise regression was used on this occasion to find the best measurement model and a stepwise regression automatically excludes any non-significant predictors.

The second part of this study correlated first year critical thinking disposition scores with the only third year dispositions that showed a significant predictive validity in 3rd year namely, the Inquisitiveness sub-scale and the SCIS composite score (combining Systematicity, Inquisitiveness and Self-confidence). Any significantly correlating 1st year dispositions were then used as the input variables in a stepwise regression analysis with the dependant variable being the 3rd year scores on the Inquisitiveness sub-scale.

7.4.3 Results

Table 7.3 shows the significant correlations between exit level scores on the critical thinking measures, degree result and third year marks. The total score of the CCTST-UK and the evaluation sub-scale score correlate significantly while the Inference sub-scale does not. The Inquisitiveness sub-scale is the only CCTDI-UK sub-scale that correlates significantly with degree result. CCTST-UK total score and CCTDI-UK Inquisitiveness were chosen for the input variables in the regressions, results of which are shown in the next table 7.4. The reasons why they were chosen and CCTST-UK Evaluation and the CCTDI-UK composite score SCIS were dropped was because the CCTST-UK Total score and Evaluation have shared items and the highest correlating variable was chosen, i.e., CCTST-UK ($r=.488$). Likewise the SCIS and Inquisitiveness measures have shared items and the Inquisitiveness measure was chosen as it had the highest correlation with degree attainment ($r=.446$).

Table 7.3: Significant correlations between degree results and critical thinking measures

		Exit CCTST-UK Total	Exit CCTST-UK Evaluation	Exit CCTDI-UK Inquisitive	Exit CCTDI-UK SCIS
Degree Average	Pearson Correlation	.488(**)	.457(**)	.446(*)	.356(*)
	Sig. (2- tailed)	.005	.010	.012	.049
	N	31	31	31	31
Level Three Mark		.555(**)	.508(**)	.415(*)	.351
		.001	.004	.020	.053
		31	31	31	31

** Correlation is significant at the 0.01 level (2-tailed).

* Correlation is significant at the 0.05 level (2-tailed).

Table 7.4 shows results of a stepwise regression showing the r square change and significance of that change for each of the significant predictors. From the two correlates chosen for regression in Table 7.3 both scores remained in the model namely, CCTST-UK total score and CCTDI-UK Inquisitiveness sub-scale. The regression model constructed from the two measures obtained in third year, predict a total of 35.6% (23.9% + 11.7%) of the variance of degree result and 39.5% (30.9% + 8.6%) of third year marks. The significant effect of Inquisitiveness is pertinent as it would be useful to know what measures in first year can predict third year levels in this critical thinking disposition. The data collected during the longitudinal analysis from cohorts 1 and 3 can provide evidence to answer this question as the tracked students have scores for dispositions in 1st year and also in 3rd year. These relationships are looked at in Tables 7.5 and 7.6.

Table 7.4: Summary of a stepwise regression showing the model of critical thinking measures that significantly predict degree attainment

Dependant /Outcome Measure	Predictor	R square change	Sig. of change
Degree Average	Exit (N=31) CCTST-UK Total	.239	.005
	Exit (N=31) CCTDI-UK Inquisitiveness	.117	.032
Level Three Marks	Exit (N=31) CCTST-UK Total	.309	.001
	Exit (N=31) CCTDI-UK Inquisitiveness	.086	.056

Table 7.5 below shows the correlations between the significant 3rd year degree predictor CCTDI-UK, the Inquisitiveness sub-scale and the 1st year scores on the CCTDI-UK sub-scales. The purpose of these correlations was to identify input variables for a stepwise regression which will identify significant predictors of 3rd year inquisitiveness.

There are three 1st year measures which correlate significantly with the exit level inquisitiveness, namely in order of strength, Inquisitiveness, the SCIS composite and self-confidence. However, Inquisitiveness and Self-confidence were chosen for the stepwise regression presented in Table 7.6 and the SCIS composite was dropped because it shared items with the other two scales.

Table 7.5: Correlations between entry dispositions and the third year levels of Inquisitiveness

		Entry Analyt .	Entry Inquis .	Entry Open- minde d	Entry Syste m	Entry Self- Confid .	Entry Truth- seekin g	Entr y TOA	Entr y SCIS
Exit Inquis .	r	-.013	.594	-.015	.114	.474	.055	.010	.516
	Sig .	.923	.000	.909	.377	.000	.674	.938	.000
	N	62	62	62	62	62	62	62	62

Table 7.6 shows the results of a stepwise regression. The only measure remaining in the model is 1st year inquisitiveness with 1st year Self-confidence being excluded because it did not significantly predict exit levels of Inquisitiveness. However, 1st year Inquisitiveness was found to be a highly significant predictor of exit level Inquisitiveness, predicting 35.2% of the variance. It is important to point out that the influence of Self-confidence cannot be completely discounted because of the moderately high inter-correlation between Self-confidence and Inquisitiveness at entry level ($r=.534$).

Table 7.6: R-square values from a stepwise regression on 1st year Inquisitiveness with the dependant variable of 3rd year Inquisitiveness

Dependant/Outcome Measure	Predictor	R square	Sig. of r-square
Exit CCTDI-UK Inquisitiveness	Entry (N=62) CCTDI-UK Inquisitiveness	.352	.000

7.5 Discussion

The first study in this chapter provides evidence of the predictive relationships between critical thinking and degree attainment. These relationships are given added ecological validity by controlling for A-levels in the analysis. In other words the analysis does provide evidence for the added benefit of using critical thinking tests for select students when used alongside A-levels. The main findings here were that a mere 8 items of the Inference sub-scale completed in 1st year were significantly predicting 3.8% of degree result and 5.6% of 3rd year marks. Furthermore, the comprehensive assessment programme associated with A-levels are only providing slightly more predictive strength, 10.4% of degree result and only 7.8% of third year marks.

The next study in this chapter shows that third year scores on the CCTST-UK and CCTDI-UK Inquisitiveness sub-scale predict 35.6% of the variance of degree result (23.9% and 11.7% respectively) and an even higher percentage, 39.5% of third year marks (30.9% and 8.6% respectively). Furthermore 1st year scores on the CCTDI-UK Inquisitiveness subscale predict 35.2% of the variance in the third year levels of the important Inquisitiveness predictor.

The following diagrams summarise the main findings of the chapter on predicting degree results (Figure 1) and third year (Figure 2) marks using critical thinking while controlling for A-levels.

Figure 7.1: Regression model based on percentage of variance of Degree result predicted by critical thinking and A-levels

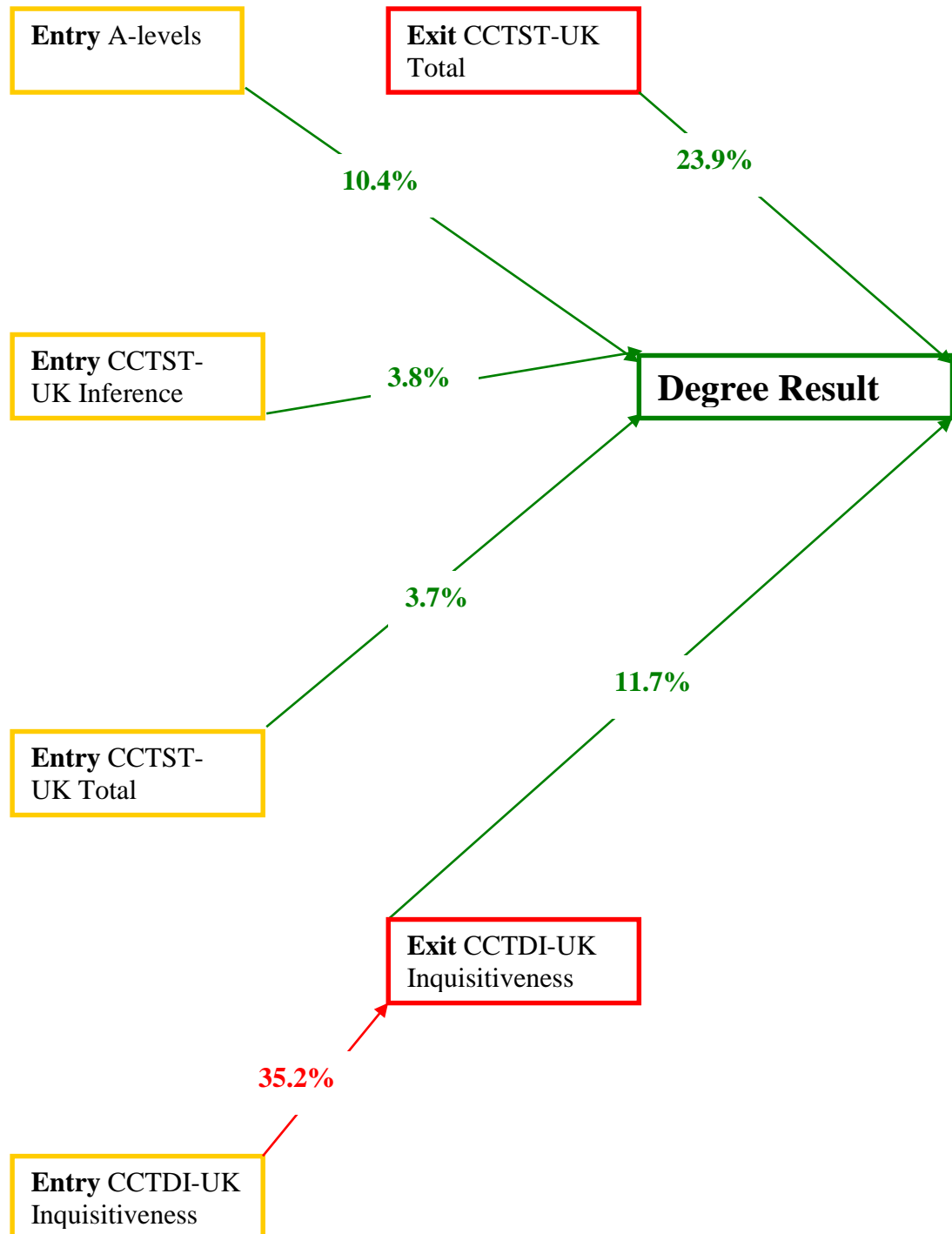
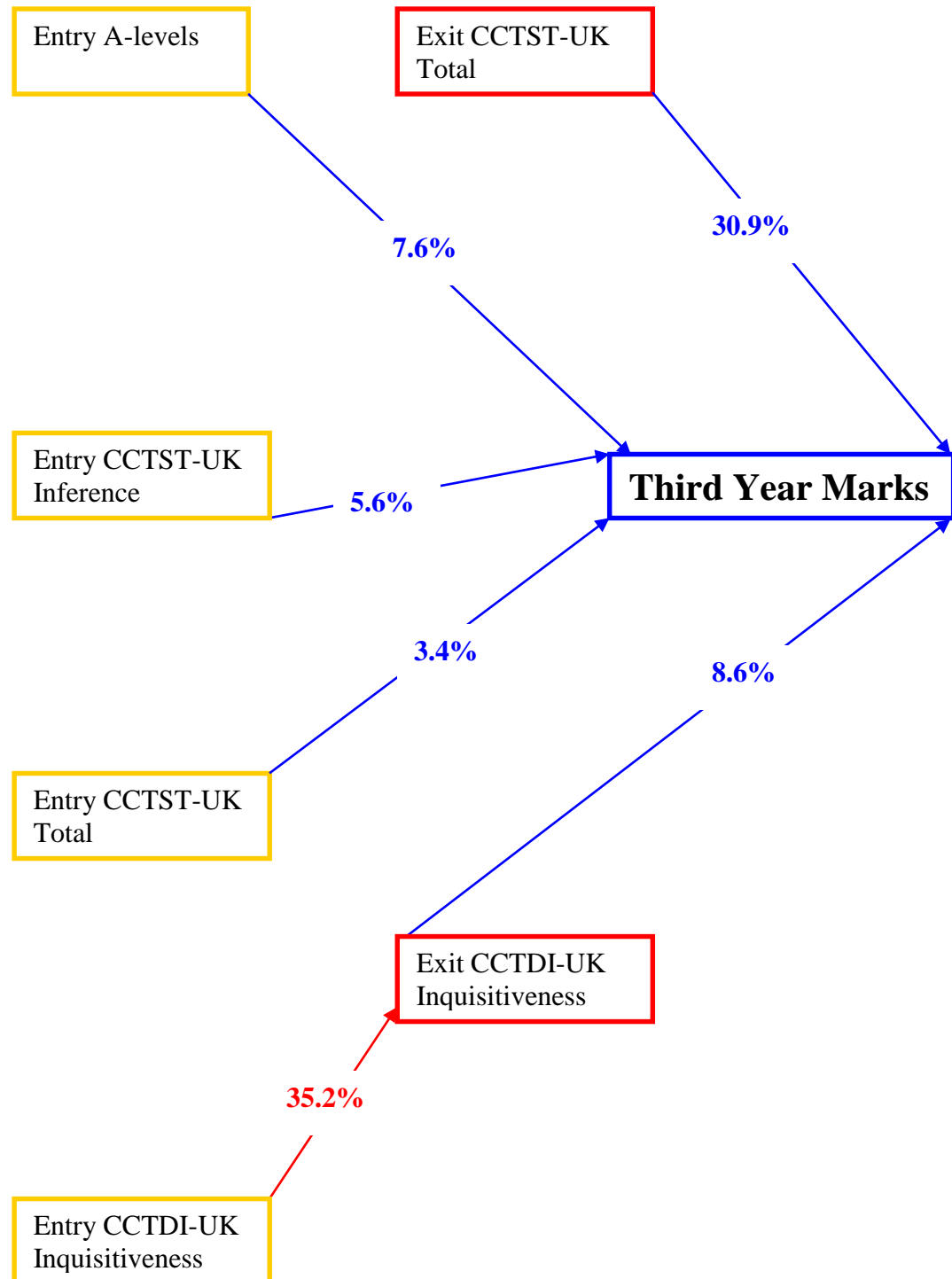


Figure 7.2: Regression model based on percentage of variance of third year degree marks predicted by Critical Thinking and A-levels



In conclusion it can be stated that the revised measures produced in chapter 5 and 6 have varying levels of predictive validity, with critical thinking skills being superior to critical thinking dispositions in this respect. Scores on the CCTST-UK sub-scale Inference and total score of the CCTST-UK were identified as being the best predictors after A-levels and the RAPM-sf did not show any significant predictive validity. Also exit scores on the CCTDI-UK sub-scale 'Inquisitiveness' showed some predictive validity.

The second aim of the chapter of this chapter also produced some interesting and applicable results in that these tests displaying predictive validity have potential to be used alongside A-levels to help universities select students for higher education. This is particularly useful in the current situation where A-levels are coming under criticism for a loss of discriminating power.

Chapter 8

General Discussion

8.1 Introduction

The final chapter is a general discussion which attempts to synthesize and critique the findings of the thesis. The chapter also proposes practical and theoretical implications of the work and suggests future directions based on the findings. The chapter is divided into five sections. The first section summarises and combines the findings from all three empirical chapters in the thesis, with the aim of producing evidence based responses to the research questions outlined in the rationale. The next section proposes a number of practical and theoretical implications of the thesis. It does this by suggesting where the work in the thesis adds to the current notions in critical thinking theory and practice. The penultimate section identifies some topics for future research, suggesting theory and methods around critical thinking measurement that could be explored. The final section is a brief conclusion on the work in the thesis.

Before proceeding with the discussion and conclusions, some cautionary points need to be raised. Firstly, the scope of the constructs related to critical thinking is limited. Even the constructs that were mapped by the Delphi Report and by the major authors in the field were not fully assessed by the instruments used in the studies. Only the critical thinking skills, Analysis, Inference and Evaluation, were assessed and, after psychometric refinement, only Inference and Evaluation remained. The scope of critical thinking dispositions was wider and initially included Inquisitiveness, Self-Confidence, Systematicity, Truth-Seeking, Open-Minded, Analyticity and Maturity; but the Maturity sub-scale was dropped due to poor reliability.

Secondly, despite the exhaustive data analyses and the progress that was made, the psychometric properties of the Californian tests for use with UK populations was limited. Whatever conclusions are drawn are constrained by that fact.

Finally, although references are made to undergraduate UK populations, the sample is narrowly based and consists of mainly psychology students in one UK university. The question of how generalisable the findings are for other groups of UK students is still open.

8.2 Synthesis of Empirical Findings

While reviewing the research questions outlined in the rationale for the thesis (Chapter 4), it was found that the research questions were answered more comprehensively when the results from a number of studies were combined rather than by examining findings from individual studies. Furthermore, the results of some studies gave an insight into more than one research question. The purpose of the following discussion is to elucidate on some of these combinations of results from different studies and how they relate to the research questions.

The first research question in the thesis was ‘What is the construct validity of the CCTDI for a UK undergraduate population?’, while the second question was ‘Is there evidence of convergent and group contrast validity (criterion validity) for the CCTDI-UK?’ and the sixth question was ‘Do critical thinking skills predict additional variance in degree outcome beyond that predicted by A-level marks?’ These may seem like independent questions but they have a shared task in that they

are assessing, in different ways, the validity of the CCTDI/CCTDI-UK. Furthermore, these forms of validity (construct, convergent, group contrast and predictive) are co-dependent and examining the overall validity profile can provide information on each of the component forms. Study 1 and Study 5 both investigate the construct validity of the CCTDI-UK but propose two different solutions; namely, the multi-variate (six factors) solution based on factorial construct validity gained from item analysis and factor analysis of individual items, and the bi-variate solution (two composite factors) gained from factor analysis of sub-scale scores. However, it is the criterion validity analysis presented in Study 4 and Study 9 that identify the ‘best’ solution by showing that the multivariate structure has superior criterion validity, and that the inquisitive subscale (from the six factor solution) shows better predictive validity than SCIS (from the 2-factor solution) Also, Figure 5.3 shows that, overall, the six disposition factors in the multi-variate solution have roughly equivalent validity levels but further scrutiny suggests that each have their own particular strengths. For example, some sub-scales are high in reliability and low in validity while others have the opposite characteristics, and some sub-scales show a more uniform pattern of reliability and validity. Reviewing the findings from several studies then would indicate that critical thinking dispositions are best investigated in their multivariate form as each scale has its own unique psychometric properties.

The next group of research questions concern the reliability and validity of the CCTST/CCTST-UK. The third research question asked ‘What is the reliability of the CCTST in UK undergraduate student populations?’ and the fourth asked ‘Is there evidence of convergent and group contrast validity for the CCTST-UK?’ While the seventh question asked, ‘Do critical thinking skills predict additional variance in

degree outcome beyond that predicted by A-level marks?’ Again when combined these three questions ask a more general question about the validity of critical thinking skills measurement. The results in Study 6 show that the remaining subscales (Evaluation and Inference) have similar internal consistency to that of the total score, although all three reliability levels are relatively low. Therefore it is unclear whether a more differentiated view of critical thinking skills (represented by subscale scores) or a more global assessment (represented by total score) is best for critical thinking skills measurement. However, the results in Study 7 show good convergent and group contrast validity for both subscales and total score. Furthermore, the results of Study 8 and 9 show good predictive validity for all three measures, i.e., the two subscales and the total score. Entry level measurement on the Evaluation subscale was a good predictor of early degree attainment (at the end of 1st and 2nd year), and Entry level assessment on the Inference subscale was shown to be a significant predictor of final degree attainment, particularly third year results. Entry level total score was also a useful predictor of degree results and finally, exit level total scores also had good predictive strength for degree marks. In conclusion, both the CCTST-UK subscales and total scores have a consistent validity profile and the subscales can probably be combined without greatly compromising the unique psychometric properties of the individual scales. Nevertheless, there may be occasions where it might be more appropriate to use the subscales rather than the composite score as revealed in Studies 8 and 9 with regard to the prediction of 1st and 2nd year exam marks compared to 3rd and overall degree mark. This conclusion is different from the conclusion drawn with regard to critical thinking dispositions.

One question which was not directly posed in the earlier chapters was about the interrelatedness of critical thinking skills and dispositions. Most theoretical frameworks reviewed made either explicit or implicit claims about the relationship between critical thinking skills and dispositions, implying that there should be a positive relationship. Nevertheless, the nature and complexity of the relationship has rarely been statistically explored. Limited research by N. Facione (1997) indicates that linear relationships may be weak and more complex developmental patterns may need to be explored. The data set from the current thesis is limited in fully investigating the relationship, as skills and dispositions were not measured for all cohorts in the sample. However, the analyses in Study 8 and 9 offer some insight into this question in that they show that critical thinking skills and dispositions both have shared potential for predicting degree attainment. Table 8.1 below explicitly shows the relationship between scores on the two critical thinking tests. These correlations are on the test scores from the students in Cohort 2 and Cohort 4 who completed both the CCTDI-UK and the CCTST-UK (N=150). Out of a possible 18 correlations, only 5 are statistically significant. This table shows that scores on the Truth-Seeking and Self-Confidence sub-scales from the CCTDI-UK are positively correlated with the CCTST-UK scores. The Truth-Seeking subscale shows the highest correlations and the pattern is similar for both the Evaluation and Inference sub-scales and for total score. For the Self-Confidence, the correlations reach significance only for the total score and for the Evaluation sub-scale.

Table 8.1: Correlations between critical thinking dispositions and critical thinking skills

		CCTDI-UK Analy.	CCTDI-UK Inquis.	CCTDI-UK Open.	CCTDI-UK Sys.	CCTDI-UK SeCon.	CCTDI-UK TrSe.
CCTST-UK Total Score	r	.130	.097	.011	.012	.180(*)	.293(**)
	Sig. (2-tailed)	.113	.237	.894	.883	.028	.000
	N	150	150	150	150	150	150
CCTST-UK Evaluation	r	.130	.040	.002	.024	.189(*)	.245(**)
	Sig. (2-tailed)	.114	.628	.984	.770	.021	.003
	N	150	150	150	150	150	150
CCTST-UK Inference	r	.087	.131	.022	.006	.109	.250(**)
	Sig. (2-tailed)	.289	.109	.789	.943	.183	.002
	N	150	150	150	150	150	150

** Correlation is significant at the 0.01 level (2-tailed).

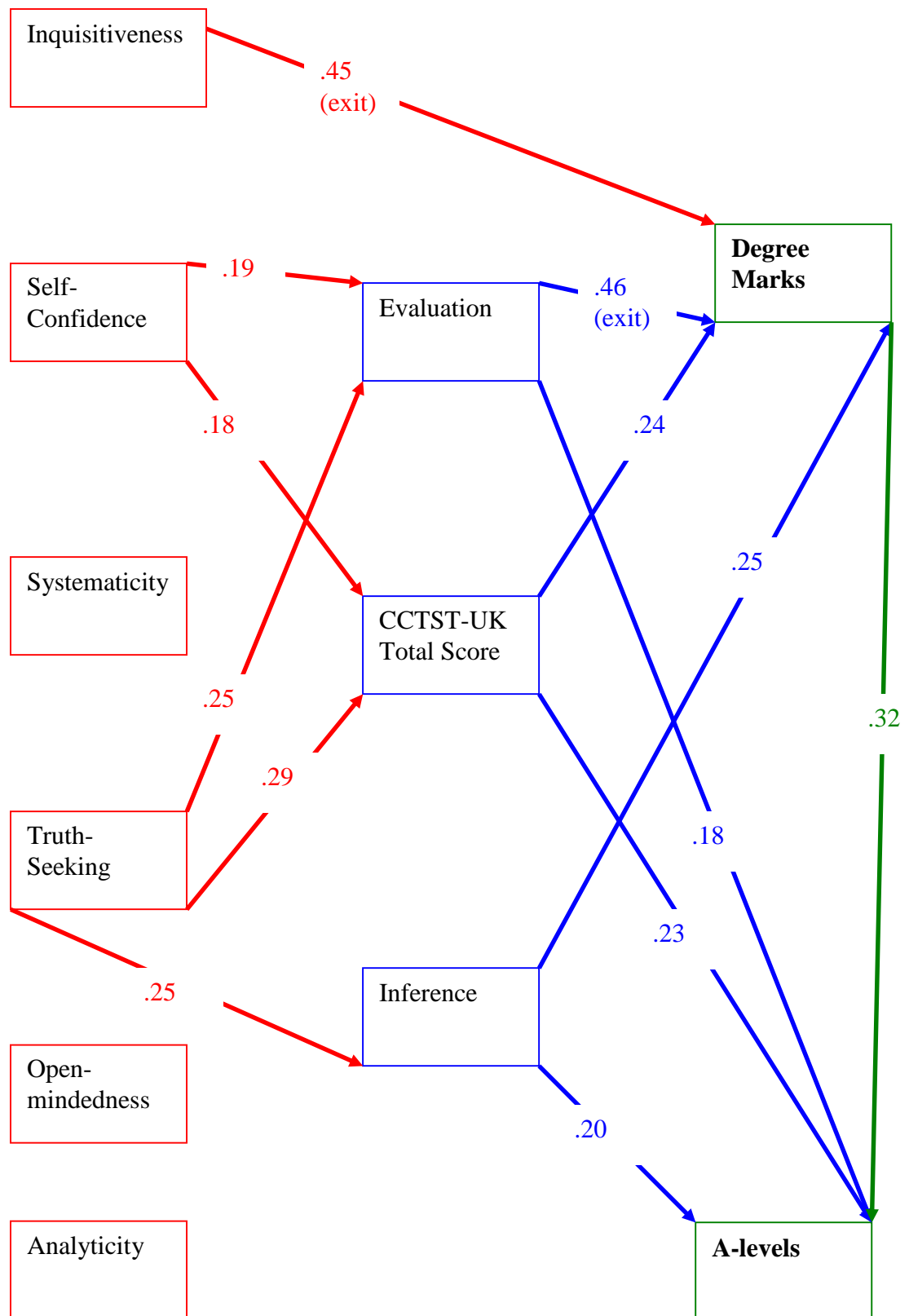
* Correlation is significant at the 0.05 level (2-tailed).

At first sight, these correlations might suggest that the interrelationships between the two types of critical thinking components are slight. However, when these relationships are considered alongside previous findings in Table 5.18, and 7.1, the participation of dispositions in critical thinking appears more prominent. The correlations in Figure 8.1 are based on first year students scores on the critical thinking measures except for two instances, namely between Inquisitiveness, Evaluation and Degree Marks where exit results have been used.

Figure 8.1 shows that the most prominent relationships occur between critical thinking skills and degree attainment. There are some direct relationships between measures of critical thinking skills and dispositions, and at least one disposition directly affects degree attainment, but dispositions may indirectly influence performance in higher education through their influence on critical thinking skills.

However, in most instances the relationships between the three main constructs (critical thinking dispositions, skills and academic attainment) are relatively low.

Figure 8.1: A model of the significant correlations between critical thinking dispositions, skills and academic attainment



In conclusion, the research in this thesis has produced empirically derived models for measuring critical thinking skills and dispositions. It also shows that the proposed models of the Delphi report have some validity for measurement. Critical thinking skills and dispositions have been shown to have identifiable structure and practical use. Lastly it could be suggested that these two critical thinking components have a tangible relationship and usefulness when combined.

8.4 Theoretical and Practical Implications

The following section discusses how the information in this thesis adds to the area of critical thinking. The main implications for the work are on the structure and measurement of critical thinking skills and dispositions. The first section looks at critical thinking dispositions and how the thesis has identified and measured six dispositions with a degree of consistency and validity. The next section looks at critical thinking skills and how two factors were identified as usable measures. Also described is their utility as a combined measure. This discussion on critical thinking skills identifies that further work is required to strengthen psychometrics on this construct. The last section discusses the potential for using critical thinking skills and dispositions in the selection of students into higher education.

The first point to be stated in this section is that the two construct model of critical thinking featuring skills and dispositions proposed by the Delphi Report (1990) has been shown to have some validity. The relationships between these two constructs explored in section 8.2 show provide some evidence of this. Also the combined utility of critical thinking skills and dispositions as predictors of degree performance provide further evidence.

Focusing solely on critical thinking dispositions, the thesis raised a number of theoretical points on their structure and measurement. The first major point is that although the CCTDI was criticized for being a psychometrically poor instrument it was restructured to form a relatively good measurement tool generally displaying good reliability and validity. A multivariate (six

dispositions) and bi-variate (two composite dispositions) explanation were offered for the measurement of critical thinking dispositions. Of these solutions the multivariate showed the best psychometric validity. When one reflects on two previous models put forward for critical thinking dispositions, i.e, the Delphi Report and the major authors' framework amalgam seen in Figure 2.1 a number of comparisons can be made. The number of dispositions measured by the CCTDI-UK falls between the numbers offered by both these other measurement models. The Delphi Report offers nineteen dispositions yet the test developed out of that report, the CCTDI, only measures seven. Whereas the revised version produced in this thesis, the CCTDI-UK, measures six of those seven dispositions, with maturity being the only one not measured. The Figure 2.1 shows only four disposition groups truth-seeking, social awareness, honesty and open-mindedness. Open-mindedness and Truth-seeking are measured in the CCTDI-UK. However it could be argued that social awareness and honesty, which may be equated to the missing sub-scale from the CCTDI, Maturity, are not being measured by the CCTDI-UK. The theoretical implications of this are discussed later in this section.

The thesis also gives an insight into the structure and measurement of critical thinking skills. The CCTST-UK provides a two component measurement model of critical thinking skills featuring Evaluation and Inference. Furthermore, the combination of scores on these measures also shows good criterion validity. Therefore the construct composed of Evaluation and Inference is possibly a secondary level skill that is a partial representation of general critical thinking ability. The reliability of the CCTST-UK was relatively poor. However, the test did show impressive criterion validity particularly in terms of predictive validity.

The CCTST-UK scope is quite restricted compared to the theoretical structure offered by both the Delphi Report and the framework synthesis displayed in Figure 2. The Delphi Report offers several additional theoretical constructs and an additional measurement construct ‘Analysis’ as outlined in the CCTST manual (Facione, 1998). Figure 2 shows that the main theorists offer many more skills including constructs associated with dynamic utilisation of a combination of skills, i.e., integration, skills involved with communication like clarification and skills recognisable with traditional study of cognitive processes like memory and judgement. In conclusion, the critical thinking skills measured in this thesis require increased consistency of measurement and greater breadth in scope.

The thesis has also explored the relationship between critical thinking skills and dispositions both in the regression models in Chapter 7 (Figure 7.1 and 7.2) and in the correlations displayed in Figure 8.1. These diagrams do illustrate a moderate interconnection between the two major constructs of skills and dispositions. However the current selection of tests measures critical thinking skills and dispositions as separate entities. This is problematic especially when the Delphi report (1990) suggests that a skill must be used in conjunction with a disposition to exercise and develop that skill appropriately.

Chapter 7 investigated the potential of critical thinking assessment as a method of selecting people for higher education. We saw in the introduction to this chapter how there are many considerations when selecting students for higher education. Many authors also highlight the potential of critical thinking selection. The following quotation suggests that there may be reluctance to use critical thinking

aptitude tests for selection. “For various reasons, Admissions Committees are unlikely to use psychometric tests for measuring personal qualities (Powis, 1994) despite some encouraging evidence about the use of such tests for predicting the clinical performance of medical students and for predicting job performance (Shen and Comery, 1997; Jelley, Parkes and Rothstein, 2002)”. Others are more optimistic and suggest that critical thinking may become the focus of educational and employment selection (Helmstadter, 1985). It is felt that the results in Chapter 7 showing the potential of critical thinking for predicting degree performance provide further empirical evidence for using critical thinking assessment as a selection method into higher education. The current format of the CCTDI-UK and CCTST-UK may not be appropriate for this role. Low reliability, validity and lack of scope of the CCTST-UK means this test may not be suitable for high stakes selection and tenuous links between critical thinking dispositions and academic attainment also raise doubt on its effectiveness as a practical selection instrument. However, properly constructed critical thinking tests with better psychometric properties (discussed in section 8.4) may be a way forward.

There are also considerations other than the psychometric properties of a test used for selecting higher education students and these are the constructs on which these students are selected. Therefore it is crucial that care is exercised when choosing these constructs as the policy makers and universities who set and impose these criteria are the gatekeepers to the valuable resource that is higher education and these people will have a partial responsibility for directing the development of society. Firstly, the evidence suggests that critical thinking in general is a worthwhile selection criterion. Critical thinking or equivalent concepts have been

repeated as a desirable habit of mind by many great thinkers, and its development has been outlined as a central aim of higher education. The next issue is that if critical thinking selection is in widespread use and thus having an important impact on society then which critical thinking constructs are being assessed is crucially important. It can be stated that the ones identified in the two tests warrant the position due to their overlap with prior critical thinking models and their predictive strength. However there are a number of important factors missing from these tests that may be important when discussing issues effecting society. These factors have been outlined by many of the major frameworks especially in the frameworks outlined by Ennis (1996a), Paul (1993) and the Delphi report (1990). These models include disposition components such as social awareness and critical thinking skills like communication and clarification which Ennis (1996a) states are essential for a properly functioning democratic society. He goes on to explain that when clarifying and communicating critical thoughts a skilled critical thinker must care about the dignity and worth of everybody affected.

In conclusion critical thinking skills and dispositions have some validity as selection criteria, with the CCTST-UK instrument showing the most potential of the two investigated in the thesis. However, a high stakes usage of aptitude testing such as academic selection requires the evidential support of careful research.

8.5 Future Directions

The following section identifies areas of future research and also suggests a number of applications of this research. Previously, the discussion has identified a

number of problems with critical thinking measurement. Firstly, assessment of critical thinking skills needs to be improved in terms of reliability and to be widened in its scope. Also, critical thinking dispositions require strengthened in their predictive validity. Furthermore, both of the two central critical thinking constructs need a socially related factor incorporated into the scales. Lastly, some form of assessment is required to measure ability to integrate critical thinking skills and dispositions. Therefore, questions raised include: What format should critical thinking tests take in terms of response format, multiple choice, open-ended or free-response? How can critical thinking constructs which are related to societal factors be measured? Are there other critical thinking factors to be identified and what are their relationships to the current stock? How are associated domains like meta-cognition related to critical thinking and how can they be assessed? What other outcomes should critical thinking predict? This is a demanding research agenda and some promising lines will be indicated.

There are a number of ways measurement could be improved in the area of critical thinking. An initial way forward is to continue the work in the thesis by improving psychometrics in the area. The thesis and a number of previous studies have highlighted the need for instruments that can measure critical thinking in a reliable and valid manner (Walsh and Hardy, 1997; Rapps 1998 and Kakai 2003). The current results have shown that a further increase in the reliability of the CCTST-UK and an additional predictive validity of the CCTDI-UK would greatly enhance the psychometrics of critical thinking.. A possible way to improve the reliability of critical thinking skills assessment is to compose new items. New items could be written for the psychometrically derived factors of critical thinking

identified in the analysis of the CCTDI-UK and CCTST-UK, avoiding the item design faults in the original. New items could also be written to improve the scope of critical thinking skills. The psychometrically poor CCTST-UK factor of Analysis previously discovered to have negligible reliability could be retested with new items, and the Maturity subscale of CCTDI which did not withstand psychometric analysis could re-investigated.

A further way of improving the measurement of critical thinking would be to explore alternative testing formats rather than multiple choice and Likert scales. Ennis (1996b) suggests that the use of multiple choice items is problematic when assessing critical thinking. For example, Inference has been identified as a measurable critical thinking skill however, these multiple choice items are assessing the ability to recognise an inference not actually produce an inference. Production or open-ended assessment would be required to assess the skill of *generating* an inference rather than merely recognising or evaluating it. The open-ended assessment of critical thinking has already been implemented OCR examinations board who have introduced an AS level (half A-level) in critical thinking. This test features multiple choice, open-ended short answer, and essay response questions. Open-ended measurement may also be useful for investigating the critical thinking skill of integration, i.e. combining a number of critical thinking skills in a goal directed way. Dispositions could also be investigated with production items by creating scenarios that tap into the participant's critical thinking attitudes. Open-ended items may also provide an opportunity to link the measurement of critical thinking skills and dispositions. Another potentially useable format of assessment is peer and teacher rating of

critical thinking skills and dispositions. In other words people who have knowledge of a student's skills and dispositions could provide an assessment of their level of development in critical thinking.

The critique section (8.4) stated that the population sampled in the thesis, i.e., psychology students, restricts the generalization of the results. Critical thinking constructs may have a different profile in students of other disciplines. Furthermore, critical thinking may predict degree attainment at a greater or lesser level. Also student populations may have particular cognitive characteristics that differ from other groups, e.g. adults who have not entered higher education or post-graduate students.

Critical thinking could also be explored in terms of its relatedness to previously unexplored critical thinking skills and associated domains. For example, are the current measures related to other critical thinking skills identified by Ennis, i.e., communication and clarification? Furthermore, a number of the major authors have identified a meta-cognitive component to critical thinking. Therefore it is deemed important to investigate what involvement metacognition has in critical thinking and whether or not it has potential as a measurable construct.

The predictive validity of critical thinking could be further investigated on outcome measures other than degree attainment. For example, societal contribution could be the dependant variable in regression models rather than degree attainment. Another way of investigating this phenomenon could be to gather dispositional profiles of people who have achieved a high level of positive societal influence, which could be compared with entry students' profiles. In

addition, longitudinal studies could track the dispositional development of students during their degree and beyond.

In conclusion, initial steps which would improve measurement in the area are to strengthen psychometrics with new items and to expand the measurement techniques to include open-ended assessment. Furthermore, widening the populations being assessed will improve the generalisability of the cognitive representations of critical thinking. More challenging exercises include relating critical thinking to associated domains like metacognition and identifying the predictive validity of critical thinking assessment on outcomes such as societal contribution.

8.6 Conclusion

The initial impetus for the thesis was to evaluate the development of critical thinking in higher education. As it turned out, much of the work concentrated on mapping the conceptual structure of critical thinking and establishing reliable valid measurement in the area. The development of good assessment of critical thinking should have financial and research support as the identification of people with these skills and dispositions is essential in an increasingly complex and technical world (Helmstadter, 1985). The empirical identification of these skills and dispositions is still in its infancy. Philosophers have been debating the components, structures and development of critical thinking since the time of Socrates. Psychologists have both the conceptual and methodological toolkit to scientifically advance the field.

References

AHESG, (2003). *Admissions to Higher Education Steering Group. Admissions to Higher Education Review. Consultation on Key Issues Relating to Fair Admissions to Higher Education*. Retrieved from the World Wide Web 6.10.04
<http://www.admissions-review.org.uk/consultation.html>

Anderson, T.; Howe, C.; Soden R.; Halliday J and Low J. (2001). Peer interaction and the learning of critical thinking skills in further education students. *Instructional Science*, 29, 1, pp. 1-32.

Anderson T., and Soden R., (2001) Peer interaction and the learning of critical thinking skills. *Psychology of Learning and Teaching*, 1, (1) 37-40.

Astin, Alexander W. (1997). *What Matters in College?: Four Critical Years Revisited*. San Francisco: Jossey-Bass.

Atagai, R. (2004). School Reform Policy. Retrieved from the World Wide Web 1.04.04 <http://www.edthai.com/publication/0003/index.htm>

Atkinson, D. (1997). A Critical Approach to Critical Thinking. *TESOL Quarterly* 31, 71-94.

Barnes, C. A. (1992) (ed.). *Critical Thinking: Educational Imperative*. San Francisco: Jossey-Bass.

Barbanel, J. (1987). Critical Thinking for Developmental Students: An Integrated Approach. *Research and Teaching in Developmental Education*, 4(1), 7-12.

Baron, J. (1985). *Rationality and Intelligence*. New York: Cambridge University Press.

Berger, A. (1985) Review of the Watson Glaser Critical Thinking Appraisal. In Conoley, J. C., & Impara, J. C. (Eds.). *The ninth mental measurements yearbook*. Lincoln, NE: Buros Institute of Mental Measurements.

Beyer, B. K. (1985). Critical Thinking: What is it? *Social Education*. April, 270-276.

Blaine-Carpenter, C. & Doig, J. C. (1988). Assessing Critical thinking Across the Curriculum. In McMillan J. H. (ed.) *Assessing Students' Learning; New Directions for Teaching and Learning*. No.34. San Francisco: Jossey-Bass.

Bloom, B.S. (1956). *Taxonomy of educational objectives: The classification of educational goals: Handbook I, cognitive domain*. New York ; Toronto: Longmans, Green.

Bowles, K. (2000). The Relationship of Critical Thinking Skills and Reflective Judgement Skills of Baccalaureate Nursing Students. *Journal of Applied Development Psychology*, 4, 23-34.

Brabeck, M. M. (1983). Critical Thinking Skills and Reflective Judgement Development: Refining the Aims of Higher Education. *Journal of Applied Developmental Psychology*, 4, 23-34.

British Psychological Society (2003). *Regulations and reading lists for the qualifying exam 2003*. Retrieved from the World Wide Web 01.10.04. http://www.bps.org.uk/documents/Qualifying_Examination.pdf

Callahan C.M. (1995) Review of the California Critical Thinking Disposition Inventory. In Conoley, J. C., & Impara, J. C. (Eds.). (1995). *The twelfth mental measurements yearbook*. Lincoln, NE: Buros Institute of Mental Measurements.

Carrol, J.B. (1993) *Human Cognitive Abilities a Survey of the Factor- Analytic Literature*. Cambridge: Cambridge University Press.

CCEA, (2003). *Pathways: Towards a more coherent, enjoyable, motivating and relevant curriculum for young people aged 11-14. Part 1: Background, rationale and detail*. Northern Ireland: Council for Curriculum, Examinations and Assessment (CCEA).

Chaffee, J. (1997). *Thinking Critically (5th Ed)*. Boston: Houghton Mifflin.

Cochrane Collaboration (2004). The Cochrane Collaboration. Retrieved from the World Wide Web 01.10.04. <http://www.cochrane.org/>

Campbell Collaboration (2004). The Campbell Collaboration. Retrieved from the World Wide Web 01.10.04. www.campbellcollaboration.org

Conklin, Karen E. (1987). The impact of a critical thinking intervention on tenth grade students' expository writing skills. [Dissertation Abstract] *Dissertation Abstracts International*. Vol 47(9-A), Mar 1987, 3293, US: Univ Microfilms International.

Conoley, J. C., & Impara, J. C. (Eds.). (1995). *The twelfth mental measurements yearbook*. Lincoln, NE: Buros Institute of Mental Measurements.

Connor-Greene P. A. and Greene D. J. (2002). Science or Snake Oil? Teaching Critical Evaluation of "Research" Reports on the Internet. *Teaching of Psychology*. 29 (4), 321-324.

Colucciello, M. L. (1997). Critical Thinking Skills of Baccalaureate Nursing Students: A Conceptual Model for Evaluation. *Journal of Professional Nursing*, 13(4), 236-245.

Cooper, C. (1998). *Individual Differences*. London; Arnold.

Costa, P. & McCrae, R (1991). *NEO PI-R Professional Manual: Revised NEO Personality Inventory*. Florida; Psychological Assessment Resources, Inc.

Cromwell, L. S. (1992). Assessing Critical Thinking. In Barnes, C. (ed.) *Critical Thinking: Educational Imperative*. San Francisco: Jossey-Bass.

Darwin, C. (1859). *On the Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life*. New York: Gramercy.

Davidson, Bruce (1995). *Critical Thinking Education Faces the Challenge of Japan. Inquiry: Critical Thinking Across the Disciplines*. Retrieved from the World Wide Web 06.10.04.

<http://www.chss.montclair.edu/inquiry/spr95/davidson.html>

DeBono, E. (1976) *Teaching Thinking*. London: Temple-Smith.

De Bono, E. (1994). *De Bono's Thinking Course*. New York: Facts on File.

Dewey, J (1909). *How We Think*. D.C. Heath and Co.

Dewey, J. (1930). *Human Nature and Conduct*. New York: The Modern Library.

Dewey, J. (1993). *How we think: A restatement of the relation of reflective thinking to the educative process*. Boston: Houghton Mifflin Company.

DfES (2003). Department for Education and Skills (2003) *Widening Participation in Higher Education*. London: HSMO.

Elder, L. & Paul, R. (2003). Critical Thinking: Teaching Students How To Study and Learn (Part 4). *Journal of Developmental Education*. 27(1) 36-37.

Ennis, R. (1962). A concept of critical thinking. *Harvard Educational Review*. 32 (1), 81-111.

Ennis, R. (1987). A Taxonomy of Critical Thinking Dispostions and Abilities. In J.B. Baron & R. S. Sternberg (Eds). *Teaching Thinking Skills: Theory and Practice*. New York: W. H. Freeman.

Ennis, R. (1990) The rationality of rationality: Why think critically? in Page, R. (Ed.) (1989). *Philosophy of Education*. Bloomington, Illinois: Philosophy of Education Society.

Ennis, R. (1992). Assessing for Accountability. In Keefe, J. W. & Walberg, H. J. (eds.) (1992). *Teaching for thinking*. Reston, VA: National Association of Secondary School Principles.

Ennis, R. (1993). Critical Thinking Assessment. *Theory into Practice* 32, 3 179-186

Ennis, R. (1996a). *Critical Thinking*. Upper Sadle River New Jersey; Prentice Hall.

Ennis, R. (1996b). Critical Thinking Dispositions: Their Nature and Assessability. *Informal Logic* 18 2&3 165-182.

Ennis, R. & Millman J. (1985a). *Cornell Critical Thinking Test, Level X*. Pacific Grove, California: Midwest Publications.

Ennis, R. & Millman J. (1985b). *Cornell Critical Thinking Test, Level Z*. Pacific Grove, California: Midwest Publications.

Ennis, R & Weir, E (1985). *Ennis-Weir Critical Thinking Essay Test*. Pacific Grove, California: Midwest Publications.

Eysenck, M. and Keane, M, (2000). *Cognitive Psychology: A Student's Handbook* 4th Ed. Psychology Press; UK.

Facione, N. (1997). *Critical thinking assessment in nursing education programs: An aggregate Data Analysis*. Millbrae California; California Academic Press.

Facione, P. (1990). *The Delphi Report: Critical Thinking: A Statement Of Expert Consensus For Purposes Of Educational Assessment And Instruction*. Millbrae California; California Academic Press.

Facione, P. (2000). *The California Critical Thinking Skills Test 2000*. Millbrae, CA. Insight Assessment.

Facione, P. A., Facione, N. C. and Giancarlo, C. A. (2000). *The California Critical Thinking Disposition Inventory; Test Manual: 2000 Update*. Millbrae, CA: California Academic Press.

Facione, P.A., Facione, R.N., Blohm, S.W., Howard, K., Giancarlo, C. A. F. (1998). *California Critical Thinking Skills Test: Manual (Revised)*. Millbrae, CA: California Academic Press.

Facione, P.A., Facione, N.C. and Sanchez, C.A. (1994). Critical Thinking Disposition as a Measure of Clinical Judgement: The Development of the California Critical Thinking Disposition Inventory. *Journal of Nursing Education*, 33, 345-350.

Facione, P. A., Sanchez, C.A., Facione N. C. & Gainen, J. (1995). The Disposition Toward Critical Thinking. *Journal of General Education*, 44, 1-25.

Feldt, R. C. (1989). Reading Comprehension and Critical Thinking as Predictors of Course Performance. *Perceptual and Motor Skills*, 68, 642.

Field R. (25/4/2001). *John Dewey (1859-1952)*. Retrieved from the World Wide Web 06.10.04. <http://www.utm.edu/research/iep/d/dewey.html>

Fielding, R M (1997). A Socio-Cognitive Perspective on Cross-Cultural Attitudes and Practices in Creativity Development. *Australian Art Education* 20(1-2) 27-33.

Fisher, A. (2001). *Critical thinking: An introduction*. UK. Cambridge University Press.

Fisher, Robert (1990). *Teaching Children to Think*. Hemel Hempstead, Herts: Simon and Schuster Education.

Fisher, A. and Scriven, M. (1997). *Critical Thinking: Its Definition and Assessment*. Norwich; Edgepress.

French, C. & French, F. (1991). Evaluating Programmes That Claim to Teach Thinking Skills: Critical Issues and Options. In Mulcalvy, R. F., Scott, R.H.& Andrews, J. (Eds). *Enhancing Learning and Thinking*. New York; Prager Publishers.

Frisby, C. L. (1992). Construct Validity and Psychometric Properties of the Cornell Critical Thinking Skills Test (Level Z): A Contrasted Groups Analysis. *Psychological Reports*, 71, 291-303.

Frost, S. E. (1942). *Basic Teachings of the Great Philosophers*. New York, NY: Anchor

Geisinger, K. (1998). Review of the Watson-Glaser Critical Thinking Appraisal Form S. In Impara, J. C. (eds.) (1998). *The thirteenth mental measurements yearbook*. Lincoln, NE: Buros Institute of Mental Measurements.

Giancarlo, C. (1993) An exploration of cognitive strategies and dispositions in relation to ego resiliency. Unpublished Masters Degree manuscript, University of California Riverside. In Facione, P. A., Facione, N. C. and Giancarlo, C. A. (2000). *The California Critical Thinking Disposition Inventory; Test Manual: 2000 Update*. Millbrae, CA: California Academic Press.

Giancarlo, C. A. & Facione, P. A. (2001). A Look Across Four Years at The Disposition Toward Critical Thinking Among Undergraduate Students. *The Journal of General Education*. 50 (1) 29-55.

Giancarlo, C. A. & Facione, P. A. (1999). *The California Measure of Mental Motivation: Manual*. Millbrae, CA: California Academic Press.

Grant, G. E. (1988). *Teaching Critical Thinking*. New York: Praeger

Gordon, T.J. (1994) The Delphi Method. Retrieved from the World Wide Web 10.10.04. http://www.futurovenezuela.org/_curso/5-delphi.pdf

Gotlieb, A. (1999). *Socrates: The Great Philosophers*. New York: Routledge.

Greenwood A. (Ed.) (1994) *The National Assessment of College Student Learning: Identification of the Skills to be Taught, Learned and Assessed*. Washington, DC: National Center for Educational Statistics.

Halonon, J., & Cromwell, L. (Eds.). (1986). *Teaching critical thinking in psychology*. Milwaukee, WI: Alverno Productions.

Halpern, Diane F. (1984) *Thought and Knowledge: An Introduction to Critical Thinking*. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Halpern, Diane F. (1994). A National Assessment of Critical Thinking Skills in Adults: Taking Steps toward the Goal. In A. Greenwood (Ed.) *The National Assessment of College Student Learning: Identification of the Skills to be Taught, Learned and Assessed*. Washington, DC: National Center for Educational Statistics.

Halpern, Diane F. (1995) [*Psychologists Teach Critical Thinking: A Special Issue of "Teaching of Psychology"*](#). N.J.; Lawrence Erlbaum Associates, Inc.

Halpern, Diane F. (1996). *Thought and Knowledge: An Introduction to Critical Thinking 3rd Ed.* New Jersey: Lawrence Erlbaum Associates.

Halpern, Diane F. (1998). Teaching Critical Thinking for Transfer Across Domains: Dispositions, Skills, Structure Training, and Metacognitive Monitoring. *American Psychologist*. Vol. 53, No. 4, pp.449-455.

Halpern, D.F. (2002). *Thinking Critically About Critical Thinking. 4th Edition (workbook)*. Mahwah, N.J.: Lawrence Erlbaum Associates.

Hanley, Gerard L. (1995). Teaching critical thinking: Focusing on metacognitive skills and problem solving. [Journal Article] *Teaching of Psychology*. Vol 22(1), 68-72. Lawrence Erlbaum, US

HEQC. (1997). *Graduate Standards Programme Final Report: Vol. 1 The Report*. London: HEQC.

Helmstadter, G.C. (1985). Review of the Watson Glaser Critical Thinking Appraisal. In Conoley, J. C., & Impara, J. C. (Eds.). (1985). *The ninth mental measurements yearbook*. Lincoln, NE: Buros Institute of Mental Measurements.

Higgins, L.T. & Sun C.H. (2002). The development of psychological testing in China. *International Journal of Psychology*, 1 August. 37, 4, pp. 246-254(9)

Higbee, J.L. (2003). Critical Thinking and College Success. *Research and Teaching in Developmental Education*. 20(1) pp.77-80.

Higbee, J. L. & Dwinell, P.L. (1998). The Relationship Between Student Development and the Ability to Think Critically. *Research and Teaching in Developmental Education*. 14 (2), pp.93-97.

Hogan, T. P. (2003). *Psychological Testing: A Practical Introduction*. New York: John Wiley & Sons.

Hongladarom, Soraj. (1998). Humanistic Education in Today's and Tomorrow's World. *Manusya: Journal of Humanities*, Vol (1) 47-58.

HMSO, (1997). *Higher Education in the Learning Society (The Dearing Report)*. London Report of the National Committee of Inquiry into Higher Education; London HMSO.

Hughes, J. (1992). Review of the Cornell Critical Thinking Tests. In Kramer, J. & Conoley, J. (Eds.) *The Eleventh mental measurements yearbook*. Lincoln, N.E: Buros Institute of Mental Measurements.

Insight Assessment (2003) Insight assessment. Retrieved from the World Wide Web 10.10.04. <http://www.insightassessment.com/home.html>

Jacobs, S.S. (1999). The Equivalence of Forms A and B of the California Critical Thinking Skills Test. *Measurement and Evaluation in Counselling and Development*, 31, 211-222.

James P. (2002). Ideas in Practice: Fostering metaphoric Thinking. *Journal of Developmental Education*. 25(3), 26-33.

Jelley, R.B, Parkes, M.A. & Rothstein M.G. (2002). Personality perceptions of medical school applicants. *Medical Education Online*, 7:11. Retrieved from World Wide Web 06.10.04 <http://www.med-ed-online.org>

Kakai, H. (2003). Re-Examining the Factor Structure of the California Critical Thinking Disposition Inventory. *Perceptual and Motor Skills*, 96, 435-438.

Kitchener, K.S. and King, P.M. (1994). Developing Reflective Judgement: Understanding and Promoting Intellectual Growth and Critical Thinking in Adolescents and Adults.

Kitchener, K.S. and King, P.M. (1981) Reflective judgement: concepts of justifications and their relation to age and gender. *Journal of Applied Developmental Psychology*, 2(2), 89-116.

Kline, P. (1986) *A Handbook of Test Construction*. London: Methuen and Co.

Kline, P. (1994) *An Easy Guide to Factor Analysis* London: Routledge.

Kline, P. (1998). *The New Psychometrics: Science, Psychology and Measurement*. London and New York: Routledge.

Kline, P. (2000). *A Psychometrics Primer*. London and New York: Free Association Books.

Kreiter C.D.; Stansfield B.; James P.A.; Solow C. (2003) A Model for Diversity In Admissions: A Review of Issues and Methods and an Experimental Approach *Teaching and Learning in Medicine*, 1 April, vol. 15, no. 2, pp. 116-122(7)

Kremer, J. and Scully, D. (1995). *Psychology and Sport*. London: Taylor Francis.

Kuhn, D., Amsel, E., & O'Loughlin, M. (1988). *The development of scientific thinking skills*. San Diego: Academic Press.

Kuhn, D. (1989). Children and adults as intuitive scientists. *Psychological Review*, 96, 674-689.

Kuhn, D. (1991). *The skills of argument*. New York: Cambridge University Press.

Kuhn, D. (1992). Thinking as argument. *Harvard Educational Review*, 62, 155-178.

Kuhn, D. (1999). A developmental model of critical thinking. *Educational Researcher*, 28, 16-25.

Kuhn, D. (2000). Metacognitive development. *Current Directions in Psychological Science*, 9, 178-181.

Kuhn, D. & Pearsall, S. (1998). Relations between metastrategic knowledge and strategic performance. *Cognitive Development*, 13, 227-247.

Kuhn, D., & Pearsall, S. (2000). Developmental origins of scientific thinking. *Journal of Cognition and Development*, 1, 113-129.

Kuhn, D. & Phelps, E. (1982). The development of problem-solving strategies. In Reese, H. (Ed.). *Advances in Child Development and Behaviour*. New York: Academic Press.

Kuhn, D., Pennington, N. & Leadbeater, B. (1983). Adult thinking in developmental perspective: The sample case of juror reasoning. In Baltes P. & Brim O. (Eds.). *Life-span development and behaviour*, Vol. 5 (pp. 158-195). New York: Academic Press.

Kuhn, D., Schauble, L., & Garcia-Mila, M. (1992). Cross-domain development of scientific reasoning. *Cognition and Instruction*, 9, 285-332.

Kuhn, D., Shaw, V., & Felton, M. (1997). Effects of dyadic interaction on argumentative reasoning. *Cognition and Instruction*, 15, 287-315.

Kuhn, D., Weinstock, M. & Flaton, R. (1994). How well do jurors reason? Competence dimensions of individual variation in a juror reasoning task. *Psychological science*, 5, 289-296.

Lambright, L. (1995). Creating a dialogue Socratic seminars and educational reform. *Community College Journal*, 65, 30-34.

Lawson, A.E. (2004). Reasoning and brain function. In R.J. Sternberg & J.P. Leighton (Eds.), *The Nature of Reasoning*. New York: Cambridge University Press.

Lipman, M. (1987). Some thoughts on the foundation of reflective education. In Baron, J. B. and Sternberg, R. J., (Eds). *Teaching thinking: Theory and practice*, (150-161). New York: Freeman

Lipman, M., Sharp, A. & [Oscanyan](#), F. (1980). *Philosophy in the Classroom*. Philadelphia : [Temple University Press](#).

Lrynock, Karoline & Robb, Louise (1999). Problem solved: How to Coach Cognition. *Educational Leadership*, 57:29-32.

MacDonald Ross, (1993). Socrates vs. Plato. The Origins and Development of Socratic Thinking. *Aspects of Education* 49, 9-22

Malcolm, K. (1992). Review of the Cornell Critical Thinking Tests. In Kramer, J. & Conoley, J. (Eds.) *The Eleventh mental measurements yearbook*. Lincoln, N.E: Buross Institute of Mental Measurements.

Martin, J. R. (1992). Critical Thinking for a Humane World. In S. P. Norris (Ed.), *The Generalizability of Critical Thinking*. (pp. 163-180). New York Teachers College Press.

Martinsons, Maris G.& Martinsons, Aelita Brivins, (1996). Conquering cultural constraints to cultivate Chinese management creativity and innovation. *Journal of Management Development*, 15 (9), 18-35.

Matilal, Bimal Krishna. (1990). *Logic, Language and Reality: Indian Philosophy and Contemporary Issues*. Delhi: Motilal Banarsidass.

McEwen A.; McGuinness C.; Knipe D. (1999) 'Comparing teaching and learning in A levels and Advanced GNVQs'. *General Educator*, 50, 12–15.

McGuinness, C. (1999). *From thinking skills to thinking classrooms: A review and evaluation of approaches for developing pupils' thinking*. London: HMSO (DfEE Research Report 115).

McMorris, R. (1995) Review of the California Critical Thinking Skills Test. In Conoley, J. & Impara, J. (eds.) (1995). *The Twelfth mental measurements yearbook*. Lincoln, Nebraska: University of Nebraska Press.

McPeck, J. E. (1981). *Critical Thinking and Education*. Oxford; Martin Robinson.

McPeck, J. E. (1990). *Teaching Critical Thinking Dialogue and dialectic*. New York: Routledge.

Meyers, C. (1987). *Teaching Students to Think Critically: A Guide for Faculty in all Disciplines*. San Francisco: Jossey Bass.

Michael, W. (1995) Review of the California Critical Thinking Skills Test. In Conoley, J. & Impara, J. (eds.) (1995). *The Twelfth mental measurements yearbook*. Lincoln, Nebraska: University of Nebraska Press.

Moran, A. P. (1997). *Managing your own Learning at University: A Practical Guide*. Dublin; Betaprint.

Munger, D., Anderson D., Benjamin, B., Busiel C., & Paredes-Holt, B. (1999). *Researching Online (2nd Ed.)*. New York: Longman.

National Education Goals Panel (1991) The National Education Goals Report. Washington, DC: U.S. Department of Education.

NCECT, (2004) Defining Critical Thinking. Retrieved from the World Wide Web 28.09.2004. <http://www.criticalthinking.org/University/univclass/Defining.html>

Nickerson, R.S., Perkins, D.N. & Smyth, E.S. (1985). *The teaching of thinking*. Hillsdale, N.J.: Erlbaum.

Norris, S. P. (1992). Testing for the disposition to think critically. *Informal Logic*. 2 & 3, 157-164

Norris, S. P. & Ennis, R. (1989). *Evaluating Critical Thinking*. Pacific Grove, CA: Midwest Publications.

Nunally, J. C. (1978). *Psychometric Theory*. McGraw-Hill.

Ochoa S.H. (1995). Review of the California Critical Thinking Disposition Inventory. In Conoley, J. C., & Impara, J. C. (Eds.) (1995). *The twelfth mental measurements yearbook*. Lincoln, NE: Buros Institute of Mental Measurements.

Olson, D. & Asington, J. (1993). Thinking about thinking: Learning how to make statements and hold beliefs. *Educational Psychologist*, 28(1), 7-23.

Onwuegbuzie, A. J., Schwartz, R. A., & Rice, D. (2001). The Role of Critical Thinking in Educational Research Courses. Paper presented at the annual meeting of the Association for the Advancement of Educational Research, Ponte Verdra, Florida.

Oxford Psychological Press, (2000) *Critical reasoning skills series: Land & liesure reference set*. UK: Oxford Psychologists Press.

Oxford English Dictionary (1997). *Oxford Compact Dictionary and Thesaurus*. Oxford, UK: Oxford University Press.

Pascarella, E. T.; Palmer, B.; Moye, M.; & Pierson, C. T. (2001). Do Diversity Experiences Influence the Development of Critical Thinking? *Journal of College Student Development*, Vol. 42, No. 3, pp. 257-269.

Passmore, J. (1967). On Teaching to be Critical. In Peters, R. S. (Ed.). *The concept of education* (pp.192-212). London: Routledge & Kegan Paul.

Paul, R. W. (1993). *Critical Thinking: What Every Person Needs To Survive in a Rapidly Changing World*. Rohnert Park, CA: Center for Critical Thinking and Moral Critique, Sonoma State University.

Paul, R. & Elder, L. (1996). Critical Thinking: Rethinking Content as a Mode of Thinking. *Journal of Developmental Education*. 19(3), 32-33.

Paul, R. & Elder, L. (2001). *Critical Thinking: Tools for Taking Charge of Your Learning and Your Life*. Upper Saddle River, NJ: Prentice Hall.

Paul, R. & Elder, L. (2002). Critical Thinking: Teaching Students How To Study and Learn (Part 1). *Journal of Developmental Education*. 26(1), 36-37

Paul, R. & Elder L. (2003). Critical Thinking: Teaching Students How To Study and Learn (Part 3). *Journal of Developmental Education*. 26(3), 36-37.

Perkins, D. N., Jay, E. & Tishman, S. (1993). Beyond Abilities: A Dispositional theory of Thinking. *Merrill-Palmer Quarterly*, 39(1), 1-21.

Perry, W. (1970). *Forms of intellectual and ethical development in the college years*. New York: Holt, Reinhart & Winston.

Piaget, J. (1963) *The origins of intelligence in children*. New York: W.W. Norton & Company, Inc.

Pintrich, P.R. and De Groot, E.V.(1990). Motivational and Self-Regulated Learning. Components of Classroom Academic Performance. *Journal of Educational Psychology*, 82, 33-40.

Pintrich, P.R., Wolters, C. and Baxter, G. (2000) Assessing Metacognition and Self-regulated learning. In G. Schraw (Ed.) *Issues in the Measurement of Metacognition*. Lincoln, NE: Buros Institute/The University of Nebraska.

Powis, D.A. (1995). Selecting medical students. *Medical Education*, 28:441-69.

Polite V. C. & Adams A. H. (1996). Improving Critical Thinking Through Socratic Seminars *Mid-Atlantic Laboratory for student Success (LSS)* Publication Series 3 US; Temple University. Retrieved from the World Wide Web 28th September 2004

<http://www.temple.edu/lss/pdf/publications/pubs96-3.pdf>

Poteet, J. (1989). Review of the Ennis-Weir Critical Thinking Essay Test. In Conoley, J.C. & Kramer, J. (1989). *The Tenth Mental Measurements Yearbook*. Lincoln, Nebraska: University of Nebraska Press.

Rapps, Jane A. (1998). Testing a theoretical model of critical thinking and cognitive development. *Dissertation Abstracts International: Section B: the Sciences & Engineering*. Vol 59(4-B), Oct, 1588, US: University Microfilms International.

Reed, Jennifer H & Kromrey, Jeffrey D. (2001). Teaching critical thinking in a community college history course: Empirical evidence from infusing Paul's model. *College Student Journal*. Vol 35(2) Jun, 201-215. Project Innovation of Mobile, US

Ramienne V.,(2002). Assessing and Developing Students' Critical Thinking.
Psychology of Learning and Teaching. 2, (1) 17-22

Resnick, L.B. (1987). *Education and Learning to Think*. Washington, DC:
National Academy Press.

Ruppel R. (2002). The Intellectual Foundations of Critical Thinking. Retrieved
from the World Wide Web 28th September 2004

<http://www.viterbo.edu/personalpages/faculty/RRuppel/CritThink.html#1>

Sala, Sergio Della & Logie, Robert H. (1998). Dualism down the drain: Thinking
in the brain. In Logie, Robert H. & Gilhooly, Kenneth J. (Eds.). *Working memory
and thinking. Current issues in thinking & reasoning*. UK: Hove.

Scott Card, Orson (1985). *Ender's Game*. Great Britain; Century Hutchinson.

Searle J.& McHarg J. (2003). Selection for medical school: just pick the right
students and the rest is easy! *Medical Education*, May, 37, 5, pp. 458-463(6)

Sen Amartya (2001). East and West: The reach of reason *New York Review of
Books*, July 20, 2000, p. 36.

Shen H. & Comrey, A.L. (1997). Predicting medical students' academic performances by their cognitive abilities and personality characteristics. *Academic Medicine*, 72:781-86.

Shermis, S. (1999). Reflective Thought, Critical Thinking. *ERIC Digest*. November (143) p. 4

Siegal, H. (1998). *Educating Reason: Rationality, Critical Thinking and Education*. New York: Routledge.

Sternberg, R. (2003). *Cognitive Psychology*. CA. USA; Thomson/Wadsworth.

Swartz, R.F. & Parks, S. (1994). *Infusing the teaching of critical and creative thinking into elementary instruction: A lesson design handbook*. Pacific Grove, CA: Critical Thinking Press & Software.

Tabachnick, B.G. & Fidell, L.S. (1989). *Using multivariate statistics*. New York: Harper & Row.

Thouless, Robert H. (1930). *Straight and crooked thinking*. London; Pan Books

Toulmin, Stephen (1958). *The uses of argument*. UK; Cambridge University Press,

Tompkins, G. (1989). Review of the Ennis-Weir Critical Thinking Essay Test. In Conoley, J.C. & Kramer, J. (1989). *The Tenth Mental Measurements Yearbook*. Lincoln, Nebraska: University of Nebraska Press.

Tscherbatsky, F. (1962). *Buddhist Logic*. New York: Dover.

Tsui, Lisa (1999). Courses and instruction affecting critical thinking. *Research in Higher Education*, 40(2), 185-200

Tsui, L. (2000). Effects of Campus Culture on Student's Critical Thinking. *Review of Higher Education*, Vol. 23, No. 4, pp. 421-441.

Tsui, L. (2001). Faculty Attitudes and the Development of Students' Critical Thinking. *The Journal of General education*, 50, 1.

Tsui, L. (2002). Fostering Critical Thinking through Effective Pedagogy: Evidence from Four Institutional Case Studies. *Journal of Higher Education*. 73 (6), 740-763.

Tredennick, H. (1990). *Xenophon, Conversations of Socrates*. London: Penguin.

Wudka J. (9/24/1998) *What is Ockham's Razor?* Retrieved from the World Wide Web, 06.10.04.

http://phyun5.ucr.edu/~wudka/Physics7/Notes_www/node10.html

van Gelder, T.J. (2001). The Reason! Project. *The Skeptic*, 21 no.2, pp.9-12

Walker,P. & Finney, N. (1999). Skill Development and Critical Thinking in Higher Education. *Teaching in Higher Education*, Vol. 4, No. 4, 1999.

Walsh, C. M. & Hardy, R.C. (1997). Factor Structure Stability of the California Critical Thinking Disposition Inventory Across Sex and Various Students Majors. *Perceptual and Motor Skills*, 85, 1211-1228.

Walsh, C. M. & Hardy, R.C. (1999). Dispositional Differences in Critical Thinking Related to Gender and Academic Major. *Journal of Nursing Education*, 38, 149-155.

Watson, G. & Glaser E. (1964). *Watson-Glaser Critical Thinking Appraisal Manual for Versions Ym and Zm*. New York: Harcourt, Brace & World.

Watson, G. & Glaser E. (1980). *Watson-Glaser Critical Thinking Appraisal Manual Form A & B*. San Antonio: The Psychological Corporation.

Watson, G. & Glaser E. (1991). *Watson-Glaser Critical Thinking Appraisal Manual Form C*. Kent: The Psychological Corporation, Harcourt Brace Jovanovich.

Watson, G. & Glaser E. (1994). *Watson-Glaser Critical Thinking Appraisal Form S*. New York: Harcourt, Brace.

Wilks, Susan (1995). *Critical and Creative Thinking: Strategies for Classroom*

Inquiry. Portsmouth, NH: Heinemann

Wollam, Jean (1992). Equality versus Excellence: The South Korean Dilemma in Gifted Education. *Roeper Review* May, 14(4), 212-17.

Appendix 2.1: Matrix of critical thinking components proposed by major authors

Author (Year)	Ennis (1996)	Paul (1993)	Fisher (1997)	Halpern (1991)	Kuhn (1999)
Definition Provided	Yes	Yes	Yes	Yes	Definition Criteria
Perspective	Educational Philosophy	Educational Philosophy	Educational Philosophy and Psychology	Cognitive Psychology	Developmental Psychology
Skills	Explicit	Implicit	Explicit	Explicit	Implicit
Clarification	‘Clarify’		‘Critical interpretation’ ‘Clarification of meaning’	‘Problem solving’	
Analysis		‘Elements of reasoning’	‘Analysis of arguments’	‘Argument analysis’	
Evaluate	‘Basis’	‘Standards of Critical Thinking’		‘Hypothesis Testing’	‘Metastrategic Knowing’
Inference	‘Inference’	‘Standards of Critical Thinking’		‘Deductive reasoning’	
Integration	‘Integration’	‘Intellectual abilities’	‘Synthesis of consideration’		
Flexibility	‘Reflective Checklist’ (FRISCO)			‘Likelihood and uncertainty’	
Communication			‘Critical communication’ ‘Clarification of meaning’	‘Thought and Language skills’	
Critical thinking awareness					
Memory					
Judgement				‘Decision Making’	
Creativity			Creativity	‘Creative thinking’	
Opinion Awareness			‘Critical knowledge’		‘Metastrategic knowing’ &

Dispositions	Explicit	'Traits of a disciplined mind'	Considered secondary to skills		Epistemological Knowing
Truth Seeking	'Get it right'	'Perserverance' 'Courage' 'Confidence in reason'			Evaluative
Honesty	'Honest and clear'	'Integrity' 'Autonomy'			Multiplist
Social Awareness	'Care for dignity and worth'	'Humility' 'Empathy'	'Interpretation of context'		Absolutist
Open-minded	'Other points of view'	'Fair-minded'			
Metacognition	Explicit (FRISCO)			'Metacognitive framework'	Metacognitive Knowing
Focus	'Focus'			'What is the goal?'	
Reasons	'Reasons'			'What is known'	
Inference	'Inference'			'Which skills get you to your goal'	
Situation	'Situation'				
Clarity	'Clarity'				
Overview	'Overview'			'Have you reached your goal'	

Appendix 3.1: Concurrent Validity of Critical Thinking Tests

Test or Scale	Test Information Test Version with Variable	Sample Information Population Sample Size	Concurrent Validity Pearson's r Spearman's s p -value = p	Reference See main reference section.
CCTDI	Truth Seeking Openness to Exper.	Undergraduate Students (N=200)	$r = .27, p = <.001$	Facione et al. (1998)
CCTDI	Open-Mindedness Openness to Exper.	Undergraduate Students (N=200)	$r = .33, p = <.001$	Facione et al. (1998)
CCTDI	CT-Confidence Openness to Exper.	Undergraduate Students (N=200)	$r = .25, p = <.001$	Facione et al. (1998)
CCTDI	Inquisitiveness Openness to Exper.	Undergraduate Students (N=200)	$r = .37, p = <.001$	Facione et al. (1998)
CCTDI	Truth Seeking Ego Resiliency	Undergraduate Students (N=200)	$r = .41, p = <.001$	Facione et al. (1998)
CCTDI	Systematicity Ego Resiliency	Undergraduate Students (N=200)	$r = .47, p = <.001$	Facione et al. (1998)
CCTDI	Inquisitiveness Ego Resiliency	Undergraduate Students (N=200)	$r = .39, p = <.001$	Facione et al. (1998)
CCTST	Form A GRE Total	Nursing Graduate Students (N=143)	$r = .719, p = <.001$	Facione et al. (1998)
CCTST	Form A GRE Analytic	Nursing Graduate Students (N=143)	$r = .708, p = <.001$	Facione et al. (1998)
CCTST	Form A GRE Verbal	Nursing Graduate Students (N=143)	$r = .716, p = <.001$	Facione et al. (1998)
CCTST	Form A GRE Quantitative	Nursing Graduate Students (N=143)	$r = .582, p = <.001$	Facione et al. (1998)
CCTST	Form A ACT	Entry Nursing Students (N=446)	$r = .582, p = <.001$	Facione et al. (1998)
CCTST	Form A CCTDI Total	English Speaking 10 th Grade (N=193)	$r = .41, p = <.05$	Facione et al. (1998)
CCTST	Form A CCTDI Total	Entry Nursing Students (N=1557)	$r = .201, p = <.001$	Facione et al. (1998)
CCTST	Form A WGCTA	Entry Nursing Students (N=139)	$r = .405, p = <.001$	Facione et al. (1998)
CCTST	Form A SAT Verbal	Exit Nursing Students (N=65)	$r = .65, p = <.001$	Facione et al. (1998)
CCTST	Form A SAT Verbal	Undergraduate Students (N=333)	$r = .55, p = <.001$	Facione et al. (1998)
CCTST	Form A SAT Math	Entry Nursing Students	$r = .422, p = <.001$	Facione et al. (1998)
CCTST	Form A SAT Math	Undergraduate Students (N=333)	$r = .44, p = <.001$	Facione et al. (1998)
CCTST	Form A College GPA	Undergraduate Students (N=473)	$r = .20, p = <.001$	Facione et al. (1998)
CCTST	Form A Nelson-Denny	Undergraduate Students (N=42)	$r = .49, p = <.001$	Facione et al. (1998)
CCTST	Form A Age	Undergraduate Students (N=479)	$r = -.006, p = <.449$	Facione et al. (1998)
CCTST	Form A Pre-test Form A Post-test	Undergraduate Students (N=479)	$r = .70, p = <.001$	McMorris (19??)
Cornell CTT	Level Z Measures of Scholastic Aptitude	Undergraduate Students	$r = .27-.74$ Ave. $r = .50$	Hughes (19??)

Cornell CCT	Level Z Grades	Undergraduate Students	$r = .15$ and $.17$	Hughes (19??)
WGCTA	Form A&B ACT	University Entry Students (N=61)	$r = .65, p = <.001$	Watson-Glaser (1991)
WGCTA	Form A&B CAT Reading	University Entry Students (N=61)	$r = .64, p = <.001$	Watson-Glaser (1991)
WGCTA	Form A&B 1 st Term Marks	University Entry Students (N=61)	$r = .30, p = <.05$	Watson-Glaser (1991)
WGCTA	Form A&B SAT Verbal	Liberal Arts Freshman (N=60)	$r = .60, p = <.001$	Watson-Glaser (1991)
WGCTA	Form A&B SAT Math	Liberal Arts Freshman (N=60)	$r = .41, p = <.001$	Watson-Glaser (1991)
WGCTA	Form A SAT Verbal	Nursing Students (N=32)	$r = .45, p = <.001$	Watson-Glaser (1991)
WGCTA	Form A SAT Math	Nursing Students (N=32)	$r = .48, p = <.001$	Watson-Glaser (1991)
WGCTA	Form A ACT English	Nursing Students (N=24)	$r = .21, p = ns$	Watson-Glaser (1991)
WGCTA	Form A ACT Math	Nursing Students (N=24)	$r = .30, p = ns$	Watson-Glaser (1991)
WGCTA	Form A GPA	Nursing Students (N=86)	$r = .19, p = ns$	Watson-Glaser (1991)
WGCTA	Form A 1 st Term Marks	College Students (N=116)	$r = .12, p = ns$	Watson-Glaser (1991)
WGCTA	Form A SAT Verbal	College Students (N=50)	$r = .57, p = <.001$	Watson-Glaser (1991)
WGCTA	Form A SAT Math	College Students (N=50)	$r = .29, p = <.05$	Watson-Glaser (1991)

Appendix 3.2: Reliability Estimates of Critical Thinking Tests

Test or Scale	Test Information Form Sub-Scale or Overall Score	Sample Information Population Sample Size	Reliability CA=Cronbach's Alpha KR=Kuder Richardson-20 SH=Split Half TR=Test-Retest IR=Inter-Rater	Reference See main reference section.
CCTDI	Overall Score	1019 College Freshmen Students	CA = .92,.90	Facione et al. (1994)
CCTDI	Overall Score	Undergraduate Juniors	CA = .90	Walsh & Hardy, (1999)
CCTDI	Overall Score	Israel age 17-18 Italy age 18-19	CA = .71 CA = .77	Zoller et al. (2000)
CCTDI	Analyticity	1019 College Freshmen Students	CA = .72	Facione et al. (1994)
CCTDI	CT Self-Confident	1019 College Freshmen students	CA = .78	Facione et al. (1994)
CCTDI	Inquisitiveness	1019 College Freshmen Students	CA = .80	Facione et al. (1994)
CCTDI	Maturity	1019 College Freshmen Students	CA = .75	Facione et al. (1994)
CCTDI	Open-Mindedness	1019 College Freshmen Students	CA = .73	Facione et al. (1994)
CCTDI	Systematicity	1019 College Freshmen Students	CA = .74	Facione et al. (1994)
CCTDI	Truth-Seeking	1019 College Freshmen Students	CA = .71	Facione et al. (1994)
CCTST	Form A Overall Score	437 UG Students 462 UG Students 90 UG Students	KR = .69 KR = .68 KR = .70	Facione et al. (1998)
CCTST	Form A Overall Score	1461 College Freshmen	CA = .57	Jacobs (1999)
CCTST	Form A Overall Score	Nursing Students	KR = .68 - .70	Bowles (2000)
CCTST	Form A	103 Graduate Students	CA = .72	Onuegbuzie et al. (2000).
CCTST	Form B Overall Score	90 UG Students 153 Graduate Nursing Students	KR = .71 KR = .75	Facione et al. (1998)
CCTST	Form B Overall Score	1461 College Freshman	CA = .61	Jacobs (1999)
Cornell	Level X		CA = .71-.78 ave.	Ennis and Millman

CTT	Overall Score		.72	(1985a)
Cornell CCT	Overall Score	Secondary Students	KR = .72	Norris (1989)
Cornell CCT	Overall Score	College Students	CA = .70	Mines et al. (1990)
Cornell CCT	Level Z Overall Score	UG and Graduate Students	SH = .80	Frisby (1992)
Ennis-Weir	Overall Score	Students (N=27)	IR = .82	Poteet (1996)
WGCTA	Form A Overall Score	Students (N=127-212)	SH = .72-.85 ave. .80	Watson-Glaser (1991)
WGCTA	Form A Overall Score	High School Seniors and Masters Students	CA = .76	Brabeck (1983)
WGCTA	Form A Overall Score	Undergraduates	SH = .70	Feldt (1989)
WGCTA	Form A Overall Score	College Students	CA = .82	Mines et al. (1990)
WGCTA	Form A	Secondary Students	KR = .74-.92 ave. .80	Norris (1989)
WGCTA	Form A Overall Score	Pre-Service Teachers (N=123)	SH = .69-.85 TR = .73	White & Burke (1994)
WGCTA	Form A Overall Score	College Students	CA = .87	Schraw & Aplin (1998)
WGCTA	Form B Overall Score	Students (N=127-212)	SH = .70-.79 ave. .77	Watson-Glaser (1991)
WGCTA	Form B Overall Score	Undergraduates	SH = .63	Feldt (1989)
WGCTA	Form C Overall Score	Grammar School Students (N=111)	SH = .87	Watson-Glaser (1991)
WGCTA	Form C Overall Score	Police Officers (N=267)	SH = .88	Watson-Glaser (1991)